JCSCS - Journal of Computer Science and Control Systems, Vol. 5, Nr. 1, May 2012

Journal of Computer Science and Control Systems

http://electroinf.uoradea.ro/reviste/default.htm

ISSN: 1844-6043

University of Oradea Publisher

JCSCS - Journal of Computer Science and Control Systems, Vol. 5, Nr. 1, May 2012

Academy of Romanian Scientists

University of Oradea, Faculty of Electrical Engineering and Information Technology

Vol. 5, Nr. 1, May 2012

Journal of Computer Science and Control Systems

University of Oradea Publisher

Academy of Romanian Scientists
University of Oradea, Faculty of Electrical Engineering and Information Technology

Vol. 5, Nr. 1, May 2012

Journal of Computer Science and Control Systems

University of Oradea Publisher

ISSN 1844 - 6043

This volume is sponsored by *The National Authority for Scientific Research*, Romania, within the frame of Grant no. 17M/08.05.2012.

CONTENTS

BAKHT Humayun - Taitec College Manchester, United Kingdom
On-Demand Data Delivery Routing for Mobile Ad-hoc and Wireless Sensor's Network 5

BEN JOUIDA Haythem[1], LAKHOUA Mohamed Najeh[2] - [1]University of Tunis, ESSTT, Tunisie, University of Carthage, ESTI, Tunisie
Combination of the Methods OOPP and FMECA for the Analysis of Process Systems.....9

BENȚIA Ioana, SZABÓ Loránd, RUBA Mircea - Technical University of Cluj-Napoca, Romania
A Novel Rotary-Linear Switched Reluctance Motor.....13

CHINDRIS Virgil, SZASZ Csaba – Technical University of Cluj-Napoca, Romania
Real-Time Simulation Environment for Embryonic Networks 17

COSTEA Claudiu Raul – University of Oradea, Romania
A Control Design for Grinding Systems with Feedforward Compensation 23

DACHIN Tudor[1], MEZA Serban[2], NEMES Marian [3], VODA Adriana[4], BADILA Florin[5] - [1]"Lucian Blaga" University of Sibiu, Romania, [2]Technical University of Cluj-Napoca, Romania, [3]Continental Automotive Systems S.R.L, Sibiu, Romania, [4]Artsoft Cluj-Napoca, Romania, [5]Wenglor Electronic S.R.L. Sibiu, Romania
Novel current monitoring techniques without shunt resistors.....27

DAVE Dhiren, NALBALWAR Sanjay, GHATOL Ashok - Dr. Babasaheb Ambedkar Technological University, Lonere, India
Location Aware Control — A Merchant Shipping Perspective 31

HROZEK František, IVANČÁK Peter - Technical University of Košice, Slovak Republic
Depth Map Calculation for Autostereoscopic 3D Display 37

KHAN Muhammad Zahid[1], ASIM Muhammad, KHAN Ijaz Muhammad[2] - [1]Liverpool John Moores University, Liverpool, United Kingdom, [2]Dhofar University, Salalah, Oman
An Overview of Hierarchical Schemes for Fault Management in Wireless Sensor Networks 43

LUPU Ciprian, PETRESCU Cătălin – University POLITEHNICA of Bucharest
Optimization Solution for Multiple Model Control Structures 49

MARGINEAN Ana-Maria, MARGINEAN Calin, TRIFA Viorel - Technical University of Cluj Napoca, Romania
Simulation of Temperature Control in Fermentation Bioreactor for Ethanol Production.....55

PETRUS Vlad[1, 2], POP Adrian-Cornel[1, 2], GYSELINCK Johan[2], MARTIS Claudia[1], IANCU Vasile[1] - [1]Technical University of Cluj-Napoca, Romania, [2]Université Libre de Bruxelles, Belgium
Average torque control of an 8/6 Switched Reluctance Machine for Electric Vehicle Traction 59

POP Adrian-Cornel[1, 2], PETRUS Vlad[1, 2], GYSELINCK Johan[2], MARTIS Claudia[1], IANCU Vasile[1] - [1]Technical University of Cluj-Napoca, Romania, [2]Université Libre de Bruxelles, Belgium
Finite Element Based Multiphysics Optimal Design of Switched Reluctance Motors Used in Electric Vehicles Propulsion....65

POPA Dan-Cristian, GLIGA Vasile-Ioan, SZABÓ Loránd, IANCU Vasile - Technical University of Cluj-Napoca, Romania
Analytical Analysis of the Tubular Transverse Flux Reluctance Motor....71

PORURAN Sivakumar[1], MARIMUTHU Rajaram[2] - [1]SKP Engineering College Thiruvannmalai, India, [2]Anna University of Technology, Thirunelveli, India
Performance Improvement of DFE on CDMA channel....75

RAGAB Khaled - King Faisal University, Hofuf, Saudi Arabia
Peer-to-Peer Overlay Network for On-demand Video Streaming....79

RUSU Călin[1], BARA Alexandru[2], SZOKE Eniko[1], DALE Sanda[2] – [1]Technical University of Cluj-Napoca, [2]University of Oradea, Romania
Fuzzy Based Reactive Controller for a Small Mobile Robot Platform....89

SOMEŞAN Liviu, PĂDURARIU Emil, VIOREL Ioan-Adrian, SZABO Lorand - Technical University of Cluj-Napoca, Romania
Steady State and Dynamic Behavior of a Permanent Magnet Flux-Switching Machine....95

STOJANOVIC Igor[1], STANIMIROVIC Predrag[2], MILADINOVIC Marko[2], STOJANOVIC Dragana[1] - [1]'Goce Delcev' University, Stip, Macedonia, [2]University of Nis, Serbia
Application of Non-Iterative Method in Image Deblurring....99

SZILÁGYI Szabolcs[1], ALMÁSI Béla[2], - [1]University of Oradea, Romania, [2]University of Debrecen, Hungary
A Review of Congestion Management Algorithms on Cisco Routers....103

UNHAUZER Attila, VÁRADINÉ SZARKA Angéla - University of Miskolc, Hungary
Online Software Module for Measurement of Audio-Frequency Controlled Heat- Storage Power....108

YILDIRIM Şahin, ARSLAN Erdem - Erciyes University, Turkey
Design and Dynamic Analysis of Six Legged Walking Robot....112

On-Demand Data Delivery Routing for Mobile Ad-hoc and Wireless Sensor's Network

BAKHT Humayun
Taitec College Manchester
202-208 Cheetham Hill Road, M8 8LW Manchester, United Kingdom,
E-Mail:humayunbakht@yahoo.co.uk/hbakht@taitecmanchster.com

Abstract **– *Mobile Ad-hoc On-Demand Data Delivery Protocol (MAODDP) is an on-demand data delivery type protocol for mobile ad-hoc network. Protocols belong to this type have a common feature of route establishment and data delivery simultaneously one after the other. The contribution of this work is to extend MAODDP operational structure for supporting group communication in a Wireless Sensor's Network. In addition, an evaluation study of MAODD in a varying simulation environment is also presented. Evaluation results indicate that MAODDP showed an impressive performance with a good data delivery and higher memory conservation in all the conducted simulation experiments.***

Keywords: ***MAODDP; Routing Protocols; SWANS; WSN's***

I. INTRODUCTION

Routing in mobile ad-hoc network is achieved through the mutual cooperation of mobile devices that form route in between two mobile stations. It is one of the challenging issues in mobile ad-hoc network [1]. The current protocols for an ad-hoc network can generally be categorized into two groups i.e. pro-active and re-active [15, 23]. Pro-active protocols by continuously evaluating the known and attempting to discover new routes, try to maintain the most up-to-date view of the network [2]. This allows them to efficiently forward packets as the route is known in advanced [14]. In contrast reactive protocols determine the route only when require [3, 5, 6].

Mobile Ad-hoc On Demand Data Delivery protocol (MAODDP) support both unicast and multicast routing. MAODDP belong to on-demand data delivery protocol type for mobile ad-hoc network. MAODDP establishes route on demand and delivers the data at the same time one after the other [22]. It is designed to minimize reaction to topological changes. It uses combination of sequence numbers and broadcast ID to ensure the freshness of routes. MAODDP is loop-free, self-starting protocol which can scales to different size of networks. MAODDP offers quick adaptation to the dynamic link conditions with low memory overhead.

In addition to a simulation based evaluation of MAODDP, this paper presents an extended model of MAODDP which incorporates protocol support of group communication in a Wireless sensor's network (WSN). In evaluation phase of this work, we are more focused in extending previous reported research [21] by evaluating MAODDP in a different simulation environment. Rest of this paper has been organized as follows. In section 2 an introduction to MAODDP with added feature to support group communication in WSN's is presented. In section 3 simulation environment is explained. In section 4 evaluation results and observations are discussed. Conclusions and future work are given in section 6.

II. MOBILE AD-HOC ON-DEMAND DATA DELIVERY PROTOCOL (MAODDP)

Mobile Ad-hoc On Demand Data Delivery Protocol adopts an intermediate approach in between tables driven and existing on-demand routing protocols [4]. MAODDP handles routing along with some others related issues. MAODDP defines different functions; definition and explanation of these functions are as follows.

A. Broadcasting Joining Message

Joining message is broadcasted in one of two situations. If number of mobile nodes want to form an ad-hoc network they broadcast joining message as an initial point of contact for the other nodes. In the other possibility, if a mobile node wants to join an established ad-hoc network, it broadcasts join message which indicates joining of new node in the network.

B. Broadcasting Route Query and Data Delivery Packet (RQDD)

A node floods RQDD packet which contain both route query information and data if it wants to deliver data packet to a destination for which it has no route information available.

C. Forwarding Route Query and Data Delivery Packet (RQDD)

Whenever a RQDD reaches to a node, it takes following steps in sequence. Legitimacy of RQDD: if the same packet has arrived before. To find out the freshness of received RQDD, receiving node check the sequence number and the broadcast ID of RQDD against the information stored inside routing table. If sequence number and broadcast ID matches with the one stored inside receiver routing table, nodes take no further action and discards the packet. Updating routing information: A node updates routing information if it

finds no previous information about the received RQDD inside its routing table or if the RQDD contains the latest information about the source and the destination node. Forwarding RQDD: Receiving node takes one of the following actions, If the destination node is its next hop neighbor or it has no information about the destination, it forwards RQDD to destination or the next hop-neighbor. The other alternative is, if the node has fresh enough route to the destination, it forwards the RQDD using the available route.

D. Creating Reverse Routes

All the intermediates nodes or nodes RQDD passes through to destination form a reverse route from destination node to the RQDD source.

E. Broadcasting and Forwarding Acknowledge Message (ACK)

Destination nodes broadcast an ACK message when it receives RQDD from a source node. This ACK message is forwarded to the source node using the reverse route.

F. Managing Sequence Numbers

Each node is responsible for maintaining its own sequence number to ensure loop free routing.

G. Error Detection and Broadcasting Error Messages

In MAODDP, error detection occurs in one of the following conditions. When a node detects a link breakage for the next hop or neighboring node in an active route or it receives route error message (RER) from one of the neighbor informing a route failure, inactive nodes in one or more active routes. Finally, any suspected behavior of a node within an active route can cause a node broadcasting error message. Broadcast of these route error messages are limited to the neighboring nodes or to the mobile nodes using an establish route for data transfer.

H. Power Saving Mode

MAODDP is adaptable to existing power saving mechanisms [19] besides its own power saving scheme. Nodes are allowed to switch in between one of two states i.e. soft or sleep and/or active state. A node can switch into sleep mode in one of two conditions; if it does not hear from other node within a time limit known as Listening Time (LT), OR, if it is not a part of any active communication. Similarly, after a specific time interval known as Active Time (AT) a node can switch back to an active state. In practice, the time between the two modes are kept as minimum as possible to reduce loss of any incoming packet.

I. Multicast Joining Request

If a node wishes to join a multicast tree for which it is not a member. It can invoke multicast joining request by broadcasting a Multicast Joining query (MJR). On the other hand, if a node wants to share some information with a node part of some other multicast tree it can broadcast multicast route query and data delivery (MRQDD).

J. Multicasting

MAODDP has been implemented in a manner which allows integration of multicast operations of other schemes within the protocol structure. MAODDP multicast operation depends on the combination of flooding and formation of the multicast tree structure. It is known that flooding is suitable for high data traffic and offers lowest control overheads while tree-based routing reduces data traffic in the network but requires many control data exchanges. MAODDP focus on maintaining only those routes that are active. Expired or invalid route entries automatically deleted from the routing table. In MAODDP, multicast tree is maintained for the life of the multicast group.

K. Security

MAODDP deal security at an intermediate level. However, implementation of the protocol allows for other security mechanisms e.g. [17] to integrate within the protocol structure. MAODDP security mechanism uses trusted certificate server C, whose public keys known to all valid nodes. Two or more mobile nodes collectively can act as a trusted server. Keys are priority generated and are exchanged through mutual relationship between C and each node. Each node obtains a certificate with exactly a single key from the trusted certificate server on joining the network. The certificate details different aspect of connecting node such as node addresses, a public key and a time stamp T1 and T2. T1 defines the certificate issue time and T2 stands for the expiry time of the certificate. These certificates are authenticated and signed by the server C. The goal of communication between source and the destination is to make sure that data is reached safely at the destination. MAODDP allocated public key to all the mobile nodes at the joining expiry time. For each RQDD the receiver node extracts the public key from the certificate ‘C’ to validate the signature and to make sure that the certificate is not expired and is still valid. The same procedure is repeated in forwarding ACKs from the destination to the source node.

L. Communication in a Wireless Sesnor’s Networ (WSN’s)

An effective communication mechanism within a WSN is an interesting topic. MAODDP has been extending its communication support for WSN and thus introduces fundamental procedures to aid such operations. Joining messages as explained in section A of this paper is applied as it stands to establish a wireless sensor’s network. MAODDP introduces one of the two principles for cluster head selection, either a node with higher battery power and capability could be chosen as a cluster head. One other possibility is with the use of a counter which works as follows. A counter is run at the formation of a wireless sensor’s network. This counter specifies maximum number of sensor’s nodes in a cluster. Any node which receives it at the start could be chosen as a cluster head. Most of the functions as defined in MAODDP specification for Mobile Ad-hoc Network is fully applicable for both inter and intra cluster communication. In essence, unlike some of the previously proposed schemes for wireless sensors network, MAODDP can offer secure group communication for WSN’s.

III.SIMULATION ENVIRONMENT

The SuSE Linux 10.1 operating environment was used for all simulation experiments. In total six sets of experiments with each set comprising nine different tests were conducted. Simulation environment was generated using selection of different input parameters. Details of each of these parameters and how these were used are as follows Nodes were placed in a grid type area range of 5x5 to 30x30 within a two dimension fixed field size of 500X500 meters. In one set of experiments nodes were placed randomly with the same fixed field as described above. Nodes were selected from the range of 25 to 450 mobile nodes. All simulation starts at 10 seconds with a fixed resolution time of 60 seconds. MAODDP was evaluated both for short and long simulations run therefore simulation stop time was chosen from the range of 600 to 800 seconds. A fixed pause time of 10 seconds was used in all simulations. In some sets mobility was defined as static and for others following mobility models were used. Random Walk: In Random Walk Mobility model mobile nodes moves in turn. Random Way Point: Random Way Point model is an extension of the random walk model. In this model each node at the beginning of its turn first moves to a new position selected at random in the unit square.

Teleport Model: This was another model which was used in some of the simulation experiments. Packet loss for most of the experiments defined as default. Adding packet loss to the simulation does not really test anything new, since the simulation already have packet loss even without specifying it. Definition and explanation of conclusions drawn from the simulation results are as follows.

Data Delivery: It defines the ratio between the number of ACK's sent and broadcast RQDD. Route Formed: Defines number of new route added. Elapsed Time: It defines the time period in between simulation start and stop time. Memory Saved: It is the difference of total memory and memory used in a simulation cycle. Total Memory: It is the memory allocated by the SWANS based on the input parameters used in a simulation cycle. Memory Used: It defines the amount of memory used in a simulation cycle.

A.Evaluation Results

In the light of evaluation results it can be drawn that MAODDP is well suited in different types of environments. Almost all the operations as defined in MAODDP specification are practically applicable and can produce good results. Variations of different types of tests have been conducted over MAODDP. These simulation environments were selected in a manner which can best reflect the nature of communication pattern of mobile ad-hoc network. Messages activities both in terms of broadcasting RQDD's and sending ACK's were quite high as shown in Figure 1 and Figure 2 respectively. A further increase in messages activities with the addition of mobile nodes can also be observed in the same graphs. This indicates that MAODDP has a good scalability factor. It was noted that the addition of mobile nodes yields little affect on the protocol behavior and performance. Evaluation experiments were run with three above mentioned mobility models. In these experiments message activities were observed as low however, with a higher data delivery rate. Average collective data delivery of all the evaluation experiments found as 79 %. This indicates that MAODDP is capable of delivering high data rates under various network environments. A higher data delivery rate was observed for the number of nodes ranges from 100 to 125. A steady high data delivery afterwards can also be seen in Figure 3. With respect to the time, highest data delivery was at in the middle of the simulation time as shown in Figure 4. Variation of data delivery is partially due to the change in grid area which has it due affects in data communication. New route formation implies more active paths for data communication. In this context, MAODDP showed an impressive and normal behavior as some of the well known routing protocols. Probability of new route formation as shown in Figure 5 was increased with the addition of mobile nodes. In Figure 5 a sharp increment with respect to the number of nodes could be seen. Average number of new route formed calculated was 613 %. Energy conservation is an important aspect mobile ad-hoc network. On average MAODDP saved 67.5 % of available memory as shown in Figure 6. Teleport model in compression with random walk and random way point model conserved the highest amount of available memory.

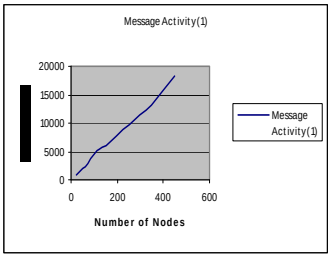

Figure 1. Messages Activities (1).

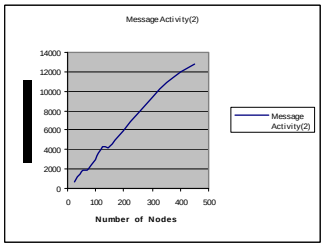

Figure 2. Messages Activities (2).

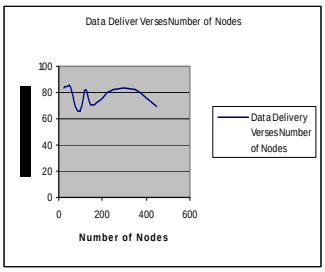

Figure 3. Data Delivery Vs Number of Nodes.

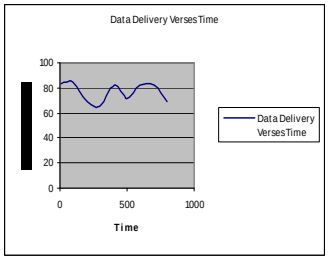

Figure 4. Data delivery Vs Time.

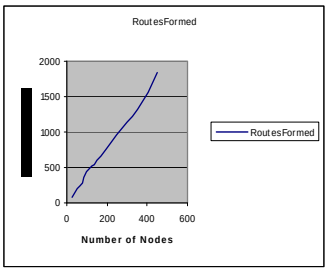

Figure 5. Routes Formed Vs No of Nodes.

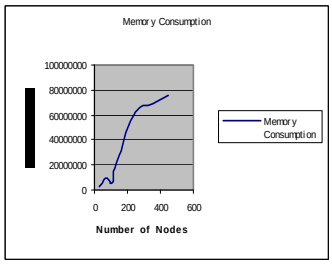

Figure 6. Memory Saved.

IV. CONCLUSIONS

This paper presents an extended model of mobile ad-hoc on-demand data delivery protocol with communication support for a wireless sensor's network. In addition, MAODDP is also evaluated in a different simulation environment. Results showed that MAODDP offers an accumulative average data deliveries of 79 % over a range of 25 to 450 mobile nodes. We are in the process of extending MAODDP basic model to evolve as a TCP based solution for routing in a mobile ad-hoc network. We are committed to share our findings with the ongoing research in this area.

REFERENCES

[1] G. Holland and N.Vaidya. Analysis of TCP Performance over Mobile Ad Hoc Networks. In Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom). 1999.

[2] S.Kannan, John E Mellor, and D.D.Kouvatsos, Investigation of routing in DSDV. 4th Annual Post-Graduate Symposium on the Convergence of Telecommunications,Networking and Broadcasting, Liverpool UK, 2003.

[3] Charles E. Perkins and Elizabeth M. Royer, Ad-hoc On-Demand Distance Vector Routing. Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, February 1999: p. pp. 90-100.

[4] H.Bakht, M.Merabti, and R.Askwith, Mobile Ad-hoc On Demand Data Delivery Protocol. 3rd Annual Post-Graduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting, 2002.

[5] H.Bakht. A Study of Routing Protocols for Mobile Ad-hoc Networks. in 1st International Computer Engineering Conference. December 2004. Cairo , Egypt.

[6] H.Bakht, M.Merabti, and R.Askwith. Centralized Frame for Routing in Mobile Ad-hoc Networks. In Proceedings of the International Conference on Computer Communication (ICCC). September, 2004. Beijing, China.

[7] D. B. Johnson and D.A. Maltz, Dynamic Source Routing in Ad-hoc Wireless Networks. Mobile Computing, 1996. 353.

[8] C. E. Perkins and E. M. Royer, Ad-hoc On- Demand Distance Vector Routing. 2nd IEEE Workshop on Mobile Computing Systems and Applications, February 1999: p. 90 100.

[9] A. Aaron and J. WengA. Aaron and J. Weng, Performance Comparison of Ad- Hoc Routing Protocols for Networks with Node Energy Constraints. June ,2001, EE 360 Class Project, Stanford University.

[10] I. D. Aron and S. K.S.Gupta, On the Scalability of On-Demand Routing Protocols for Mobile Ad-hoc Networks: An Analytical Study. Journal of Interconnection Networks, January 2001.

[11] J. Broch, et al., A Performance Comparsion of Multi-Hop Wireless Ad- Hoc Network Routing Protocols. In Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom), 1998.

[12] M.Asim and A.J. PullinM.Asim and A.J. Pullin, Comparison analysis of MAODDP with some other prominent Wireless ad hoc routing protocols, in IBITE Computing. 2005, Liverpool Hope University: Liverpool.

[13] C. E. Perkins. Performance Comparsion of two On-Demand Routing Protocols for Ad Hoc Networks. in IEEE Conference on Computer Communications (INFOCOM). March 2000.

[14] E. Royer and C.Toh, A Review of Current Routing Protocols for Ad-Hoc Mobile Wireless Networks. IEEE Personal Communications Magazine, April 1999: p. 46-55.

[15] S. Sesay, et al., Simulation Comparison of Four Wireless Ad hoc Routing Protocols. Information Technology, 2004. 3(3): p. 219-226.

[16] C.K. Toh, S.J.Lee, and M.Gerla, A simulation study of table driven and ondemand routing protocols for mobile ad hoc networks. IEEE Network, 1999. 13: p. 48-54.

[17] W. Wang, Y. Lu, and B. Bhargava. On Security Study of Two Distance Vector Routing Protocols for Mobile Ad Hoc Networks. in IEEE International Conference on Pervasive Computing and Communications. March 2003. Dallas-Fort Worth, Texas,USA.

[18] H.Bakht. A comparison based overview of destination distance sequence vector routing (DSDV) and mobile ad hoc on demand data delivery protocol "(MAODDP)",. In International Workshop on Wireless Ad-hoc Networks. May, 2005. London, United Kingdom.

[19] D. Kim, et al., Routing Mechanisms for Mobile Ad Hoc Networks based on the Energy Drain Rate. IEEE Transactions on Mobile Computing, April-June 2003. 2(2): p. 161-173.

[20] Swans, Java in Simulation Time / Scalable Wireless Ad hoc Network Simulator http://jist.ece.cornell.edu/.

[21] R.L. Gilaberte and L.P. Herrero Routing Protocols in Mobile Ad-Hoc Networks. Communication Systems and Networks – 2005.

[22] C.K. Toh, S.J.Lee, and M.Gerla, A simulation study of Tables Driven and On-Demand Routing Protocols for Mobile Ad-hoc Networks, IEEE Network, 1999. 13: p. 48-54

[23] R.L.Gilaberte, L.P.Herrero. A Secure Routing Protocol for Mobile Ad-hoc Network Based on Trust. International Conference on Networking and Services (ICNS '07). June 2007.

[24] C. E. Perkins and E. M. Royer, , Ad-hoc On- Demand Distance Vector Routing, 2nd IEEE Workshop on Mobile Computing Systems and Applications.

[25] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989. M Saleemand, MA Sheikh, An Emprical Analaysis of Ad-Hoc Routing Protocols for Hybrid Wireless Sensors Networks, International Conference on Electrical Engineering, ICEE 07, 2007.

Combination of the Methods OOPP and FMECA for the Analysis of Process Systems

BEN JOUIDA Haythem[1], LAKHOUA Mohamed Najeh[2]
[1] *Laboratory of Analysis and Command of Systems,*
University of Tunis, ESSTT
haythem.b.jouida@gmail.com
[2] *Laboratory of Analysis and Command of Systems,*
University of Carthage, ESTI
MohamedNajeh.Lakhoua@ieee.org

Abstract ***– Of the fact of the complexity of the modeling of process systems on the one hand, and the majority of methods of analysis and conception concentrate on the treatment of information on the other hand, it is necessary to adopt a structured methodology of development. The object of this paper is to propose a gait based on the combination of the two methods OOPP (Objective Oriented Project Planning) and FMECA (Failure mode, effects and criticality analysis) for the process systems analysis.***

Keywords: ***process system, analysis, OOPP, FMECA.***

I. INTRODUCTION

Today enterprises operate in a market in constant evolution, characterized by product life cycles more and shorter, an increased demand of flexibility and frequent changes of techniques and technologies.

In fact, the modeling is indispensable for the understanding and the analysis of all phenomena of the process systems. Such system conduct also rests on the utilization of models [1]. These models must give into account the structure and the behavior of the system and must permit the analysis of its qualitative and quantitative properties.

Of fact of the complexity of a process system and the interdependence of its various functions, its analysis and its conception cannot be achieved only according to a global approach [2]. This is why the gait proposed to lead such an action is a systemic gait. A systemic approach, participative is nature, is necessary to facilitate the modeling of process systems.

The risk analysis is an indispensable tool to the good conception of a device and the assessment of dangers and induced risks by its utilization. The analysis of risks takes all its value in processes of risk management developed by manufacturers [3].

The risk analysis of process systems is defined as being an analysis of the failing arborescence.

Several methods are applicable to the analysis of risk of a process system [4]: a method defined by a method for the analysis of risk and management of risks, a method for systems of information, a method defined by a guide of risk management for systems of information technology, and another guide of security management that is defined as being a guide of management of security risk. All these methods define an assessment of risks and really provide a detailed guide for the analysis of risk [5].

Other methodologies for the analysis of risk of a process system are defines by a cognitive model that puts the accent on the analysis of relations between factors of risk and risks a graphic representation feels this method.

In this perspective, a previous consists in studying the systemic analysis and risk management proposed in the literature and that will be kept in this work.

This paper is organized as follows: Section 1 presents the process systems. Section 2 gives a general presentation of the methods OOPP and FMECA. Section 3 puts the accent on a model for combination of the methods for the analysis of process systems. Finally, section 4 presents a conclusion and some perspectives of this research work.

II. PRESENTATION OF PROCESS SYSTEMS

An enterprise is an economic and social structure that regroups the human means, materials and financiers. Indeed, the enterprise combines and remunerates some necessary production factors to the creation of possessions or services. The enterprise searches for the productive efficiency that is the most efficient productive combination [6].

This efficiency of the productive combination is measured thanks to the fruitfulness. The objective of the enterprise is to improve its fruitfulness in order to increase its profitability [7].

A process system is a key element to increase the fruitfulness and the competitiveness of the industrial enterprises.

Indeed, a process system is all transformation of a whole of raw materials or components semi-finished or finished products while answering if need be for the customer and satisfactory of the various constraints (delay, cost, competitiveness, customer service, presentation, communication...) [8].

The process system puts the accent on the notion of the production that is one of the economic activities the

more old and traditional in the humanity evolution. Several researchers define the production as being a transformation of resources belonging to a productive system and driving to the creation of possessions and services. Resources can be facilities, of men, of matters (raw materials and components), of the technical information or procedures (ranges, nomenclatures, operative cards...) [9].

The process system can be modeled in three subsystems:

- The conception system that conceives some new products, modify and improve products already made and conceives manufacture tools.
- The management system that permits the management of the production, the organization and the management of stocks.
- The manufacture system that manufactures the product from the well stocked data by the conception subsystem (documents of manufacture).

In order to model a production system, it is important to have a good management of the production that is destined to apply methods and techniques in the goal to accomplish the transformation of matters in finished product.

It is the material means combination, human and of raw materials to assure the manufacture of products in quality and in quantity.

The production process makes part of a coherent chain permitting to assure the satisfaction of the customer and a profit in the enterprise. This process can be described through the following stages [10]:

- Survey of the external environmental choices, existing products and demands.
- Definition of customer needs in relation to the existing product.
- Survey of the product, its cost, the quality and the quantity asked to assure needs of customers.
- Realization of the product by the identification of production means, the determination of raw materials and the implantation of means.
- Manufacture with its phase of production and its phase of control.
- Insurance of the service after the sale.

The fast evolution of the process system in enterprises is, of the fact, that the system of production or lines of production have been automated entirely. This evolution has been advanced to the year 2000, by the existence of the harmonization of the human operators and the automatic machines what should be a human process system [11-13].

Indeed, an automated process system is a means to assure the objective primordial of an enterprise and the competitiveness of its products. It permits to add a value to the incoming products.

The notion of automated system can apply to a freestanding machine as well that to a unit of production, or even to a factory or a group of factories. It is therefore indispensable, before all analysis, to define the border permitting to isolate the automated system studied of its outside.

III. METHODS USED IN THIS RESEARCH

In this part, we present the methods OOPP and FMECA used in this research.

A. OOPP method

The OOPP method also referred to as Logical Framework Approach (LFA) [14-16], is a structured meeting process. This approach seeks to identify the major current problems using cause-effect analysis and search for the best strategy to alleviate those identified problems. OOPP method has become the standard for the International Development Project Design. Team Technologies have continued to refine the approach into TeamUP [17].

The design methodology of the OOPP method is a rigorous process, which if used as intended by the creators will impose a logical discipline on the project design team. If the process is used with integrity the result will be a high quality project design. The method is not without its limitations, but most of these can be avoided with careful use of ancillary techniques. Many things can go wrong in the implementation phase of a project, but if the design is flawed, implementation starts with a severe handicap.

The first few steps of the LFA are: situation analysis; stakeholder analysis; problems analysis [15].

The stage of “Problem Analysis” therefore seeks to get consensus on the detailed aspects of the problem. The first procedure in problem analysis is brainstorming. All participants are invited to write their problem ideas on small cards. The participants may write as many cards as they wish. The participants group the cards or look for cause-effect relationship between the themes on the cards by arranging the cards to form a problem tree (Fig.1).

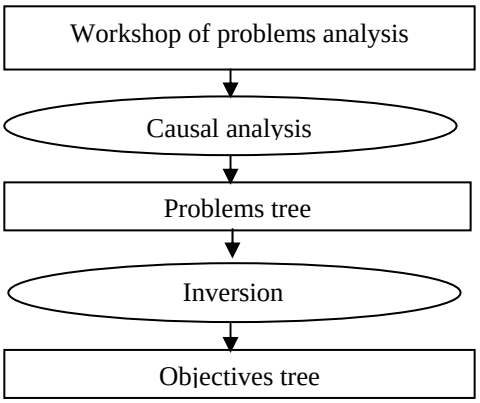

Figure 1. OOPP analysis.

In the step of “Objectives Analysis” the problem statements are converted into objective statements and if possible into an objective tree. Just as the problem tree shows cause-effect relationships, the objective tree shows means-end relationships. The means-end relationships show the means by which the project can achieve the desired ends or future desirable conditions. Frequently there are many possible areas that could be

the focus of an "intervention" or development project. The next step addresses those choices [16].

The objective tree usually shows the large number of possible strategies or means-end links that could contribute to a solution to the problem. Since there will be a limit to the resources that can be applied to the project, it is necessary for the participants to examine these alternatives and select the most promising strategy. This step is called “Alternatives Analysis”. After selection of the decision criteria, these are applied in order to select one or more means-end chains to become the set of objectives that will form the project strategy.

After defining the objectives and specifying how they will be measured (OVIs) and where and how that information will be found (MOVs) we get to the detailed planning phase: “Activities Planning”. We determine what activities are required to achieve each objective. It is tempting to say; always start at the situation analysis stage, and from there determine who are the stakeholders.

The OOPP method was extended and refined into MIDIP (Method of Specification, Development and Implementation of Project) [18-19].

B. FMECA method

The second method presented in this research is Failure mode, effects and criticality analysis (FMECA). In fact, this method is an extension of failure mode and effects analysis (FMEA). FMEA is a bottom-up, inductive analytical method which may be performed at either the functional or piece-part level. FMECA extends FMEA by including a criticality analysis, which is used to chart the probability of failure modes against the severity of their consequences [20]. The result highlights failure modes with relatively high probability and severity of consequences, allowing remedial effort to be directed where it will produce the greatest value.

The FMECA analysis procedure typically consists of the following logical steps [21]:

- Define the system;
- Define ground rules and assumptions in order to help drive the design;
- Construct system block diagrams;
- Identify failure modes;
- Analyze failure effects/causes;
- Feed results back into design process;
- Classify the failure effects by severity;
- Perform criticality calculations;
- Rank failure mode criticality;
- Determine critical items;
- Feed results back into design process;
- Identify the means of failure detection, isolation and compensation;
- Perform maintainability analysis;
- Document the analysis, summarize uncorrected design areas, identify special controls necessary to reduce failure risk;
- Make recommendations;
- Follow up on corrective action implementation and effectiveness.

Figure 2 presents a summary of the different logical steps of the FMECA analysis.

FMECA may be performed at the functional or piece part level. Functional FMECA considers the effects of failure at the functional block level, such as a power supply or an amplifier. Piece part FMECA considers the effects of individual component failures, such as resistors, transistors, microcircuits, or valves. A piece part FMECA requires far more effort, but is sometimes preferred because it necessitate more on quantitative data and less an engineering judgment than a functional FMECA [22].

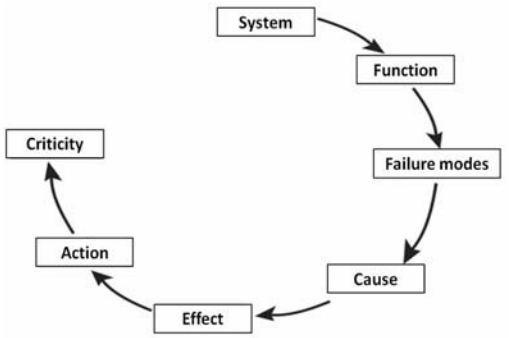

Figure 2. FMECA analysis.

The criticality analysis may be quantitative or qualitative, depending on the availability of supporting part failure data.

Strengths of FMECA include its comprehensiveness, the systematic establishment of relationships between failure causes and effects, and its ability to point out individual failure modes for corrective action in design.

Weaknesses include the extensive labour required, the large number of trivial cases considered, and inability to deal with multiple-failure scenarios or unplanned cross-system effects such as sneak circuits.

IV.MODEL OF PROCESS SYSTEMS

The diagnosis and the analysis of a process system is a complex operation. It first requires knowledge of the global working of the system and then a more and more detailed knowledge of the various components of the system in question. It is appropriated therefore to use a systemic approach that adjusts to this global gait and permitting to establish ties brings in means and ends on the basis of a previous analysis of problems, objectives and activities. This gait is based on the use of the OOPP method.

This analysis is not sufficient, it is necessary to complete it by the utilization of a method of risk management. In fact, we propose to adopt the FMECA method which is an excellent hazard analysis and risk assessment tool, but it suffers from other limitations. This alternative does not consider combined failures or typically include software and human interaction considerations. It also usually provides an optimistic estimate of reliability. Therefore, FMECA should be

used in conjunction with other analytical tools when developing reliability estimates.

Consequently, we chose to push our research on these two methods of modelling: OOPP and FMECA in order to propose a new combined modelling approach. The elaborated model was decomposed in three phases in order to assure the effective OOPP and FMECA analysis. These three phases are: OOPP analysis; FMECA analysis; Combination of the two methods in order to offer a director plan of the production.

The figure 3 presents a model of combination of the two methods OOPP and FMECA represented by its various stages.

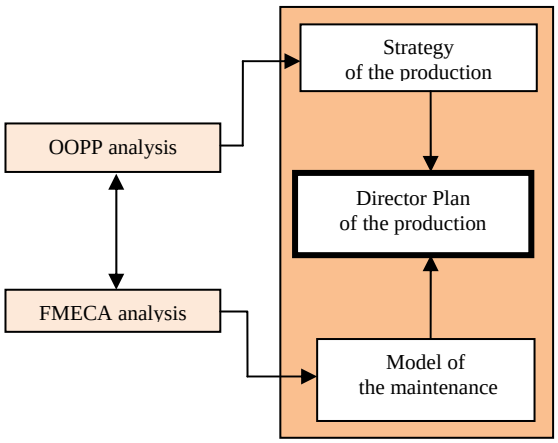

Figure 3. Model for the combination OOPP – FMECA.

The first phase according to the OOPP analysis leads to the representation of the problems tree and the objectives tree.

The analysis of the different fashions of failures modes, their effects and their criticality according to the FMECA method constitutes the second phase of the proposed model.

The combination consists to the refining of the analysis according to the two methods results.

Finally, it seems important to continue this study in order to make procedures of maintenance of the analysis and to apply the proposal of combination to practical process systems; this will give the evaluation and the improvement more easy and efficient.

V.CONCLUSION

In this paper, we presented the main methods of analysis and conception of process systems as well as the methods of risk management.

In fact, the modeling of process systems is a complex activity and requires a structured methodology of development. This is why the proposed gait is based not only on a systemic method for the representation of process systems but also on a risk management method in view of the analysis and the conception of process systems. Besides, it is necessary to adopt a method permitting to manage the risk of a process system. This method is as useful at the time of the re-conception of process systems that it was about a modernization or an optimization.

REFERENCES

[1] P. Cavelery, “Soft Systems Thinking: A Pre-condition for Organizational Learning”, Human Systems Management, p.259-267, 1994.

[2] J. Sticklen, E. William, “Functional Reasoning and Functional Modelling”, IEEE Expert: Intelligent Systems and Their Applications, 1991.

[3] A. Syalim, Y. Hori, and K. Sakurai, “Comparison of Risk Analysis Methods”: Mehari, Magerit, NIST800-30 and Microsoft’s Security Management Guide, 2009.

[4] J. Bowles, “The New SAE FMECA standard”. University of South Carolina, 1998.

[5] S. Hata, T. Anezaki, “Human Friendly Process system”, Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation, 2007.

[6] J. Ryan, C. Heavey, “Process modelling for simulation”, Computers in Industry 57, pp.437-450, 2006.

[7] A. Talbi, “Analyse de l’entreprise dans une démarche d'intégration”. JESA, 2002.

[8] P. Castagna, “Contribution à la modélisation, la simulation et la commande de systèmes de production et de transitique”, Thèse, Nantes, 2004.

[9] F. Vernadat, “Entreprise modeling and integration”. England: T.J Press Ltd, Padstow, 1996.

[10] S. Bassetto, “Contribution à la qualification et amélioration des moyens de production. Application à une usine de recherche et production de semi-conducteurs”, Thèse, ENSAM, 2005.

[11] J. Brossier, “System and farming system - A note on this concept”, 1987.

[12] N. B. Draghici, “ La modélisation et la simulation en vue de la conduite des systèmes de production”, Les Cahiers des Enseignements Francophones en Roumanie, 1998.

[13] M. Landry, “A note on the Concept of Problem”, Organizations Studies, 1995.

[14] GTZ, Methods and Instruments for Project Planning and Implementation, Eschborn, Germany, 1991.

[15] The Logical Framework Approach (LFA): Handbook for objectives-oriented planning, Norad, Fourth edition, 1999.

[16] ZOPP: An Introduction to the Method, COMIT Berlin, May 1998.

[17] S. Killich, H. Luczak, “Support of Interorganizational Cooperation via TeamUp at Internet-Based Tool for Work Groups”, Proceedings of the 6th internationally Scientific Conference, Berchtesgaden, May 22-25, Berlin, 2002.

[18] M. N. Lakhoua, “Refining the objectives oriented project planning (OOPP) into method of informational analysis by objectives”, International Journal of the Physical Sciences, Vol. 6(33), 2011, pp. 7550 - 7556.

[19] M. N. Lakhoua, T. Ben Jouida, “Refining the OOPP into Method of Representation of the Information by Objectives”, International Transactions on Systems Science and Applications, Vol. 7, No. 3/4, December 2011, pp. 295-303.

[20] L. Buzzatto, “Failure mode, effects and criticality analysis (FMECA)” Use in the federal aviation administration (FAA) reusabale, launch vehicle (RLV)” Licensing Process, 1999.

[21] T. Ben Romdhane , F. Ben Ammar , A. Badreddine, “Une approche par la logique floue pour l'optimisation multicritère de la prise de décision appliquée à l'AMDEC”, JDS, vol. 16, no4, pp. 505-544, Lavoisier, France, 2007.

[22] L. Jianping, “Study on Applying Fault Tree Analysis Based on Fuzzy Reasoning in Risk Analysis of Construction Quality” - International Conference on Risk Management & Engineering Management, 2008.

A Novel Rotary-Linear Switched Reluctance Motor

BENŢIA Ioana, SZABÓ Loránd, RUBA Mircea

Technical University of Cluj-Napoca
Electrical Machines and Drives Department
28 Memorandumului str, 400114 Cluj, Romania
e-mail: ioana.bentia@mae.utcluj.ro

Abstract ***–A novel direct-driven rotary-linear switched reluctance motor is proposed in the paper. It is able to perform both rotation and linear movement of its shaft. The motor has the advantages of robust mechanical structure, low manufacture cost and capability of operation under hostile environments. In the paper its basic structure and the working principle are presented. Advanced simulation tools were applied for checking the performances of the motor. The proposed machine is expected to be useful in diverseapplications were both precise rotary and linear motions are required.***

Keywords: ***switched reluctance motors; finite element method; rotary- linear machines.***

I. INTRODUCTION

In modern industrial environments two degrees of freedom (DoF) precise movements are frequently required. The actual solutions include rotary motors driven X-Y tables or two independent motors, one with rotary-to-linear mechanical coupling. These approaches have the disadvantages of complex mechanical structure, frequent mechanical adjustments, high manufacturing or maintenance cost and low reliability [1].

An alternative to the mechanical couplings are the rotary-linear machines which can be useful in diverse industrial and automotive applications where on a single shaft both rotational and linear movements are required.

In a direct-drive machine the mechanical energy is directly performed onto the actuator or load, thus eliminating any mechanical couplers, such as gears or belts for motion transformation. It has the advantages of fast response, high flexibility and the overall control system may have a simple structure [2].

For example in vehicles they can actuate the active-wheels or control the gearshifts in automated transmissions. Also advanced manufacturing lines require combined precise rotary and linear motion for parts assembly, drilling process, electrical wiring and component insertion, etc. [3], [4].

Until recent years switched reluctance motor (SRM) was not a popular choice for high-precision and high-speed motion applications, because it is difficult to control and its output has high torque ripples. Due to the continous advancement of power electronics and digital signal processing, and the continuous trend of simplifying the mechanics through complex control strategy a special interest was given to the switched reluctance motors.

The paper describes the development of a novel, high performance, direct-drive rotary-linear motion system. The motor is based on the variable reluctance principle. It aims to replace the traditional rotary-linear machines as a higher performance and lower cost alternative.

II. THE ROTARY-LINEAR SRM

A. Construction

The proposed electrical machine, as also its traditional rotational and linear counterparts, work upon the variable reluctance principle. Practically it is an efficient combination of a usual rotational switched reluctance machine (SRM) and a special linear SRM having several mover modules on its shaft. It has all the advantages of the SRMs: mechanical robustness, constructive simplicity, low manufacturing and maintenance costs, high reliability and relatively easy control [3].

The complex structure of the rotary-linear SRM in discussion is given in Fig. 1. As it can be seen in the figure the stator is built up modularly of three correctly shifted ordinary SRM stators having 8 poles each.

The rotor is constructed of several common 6 poles SRM rotor stacks mounted on a common shaft. The toothed structure of each rotor stack ensures appropriate flux path along the stators, air-gap and the rotor. The rotor may both rotate and move on the axial direction.

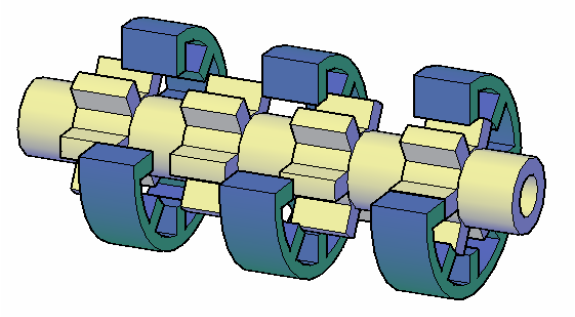

Figure 1. The structure of the proposed motor (without coils).

Each type of the movement (rotational and linear) has to be controlled independently. Only by imposing different energizing sequence for the windings on the three stators the different precise movements can be achieved [5].

The study on the two types of movements will be next described.

B. The rotation motion

When rotational movement is imposed two coils on each stator stack are fed simultaneously function of the rotor's position. The stator which has its poles aligned in the axial direction with the rotor poles will develop most of the torque. The other two stator stacks will also contribute to the rotational movement. As they are symmetrically unaligned on the axial direction in that position the axial forces developed by them will be equal but of opposite direction, hence their sum will be nil and no linear movement will be produced.

The analysis was performed for a sample machine having the following rated data:

i.) Voltage and current: $U_N = 300$ V, $I_N = 5$ A

ii.) Power: $P_N = 350$ W

iii.) Angular and linear velocity: $n_N = 600$ r/min, $v_N = 0.5$ m/s

iv.) Torque and axial force: $T_N = 5.5$ N·m, $F_N = 20$ N (for a single module).

For the study of rotary movement a two dimensional finite elements method (FEM) based numeric field analysis was performed by using the Flux 2D program. The flux lines and the coresponding flux density map are given in Figs. 2 and 3.

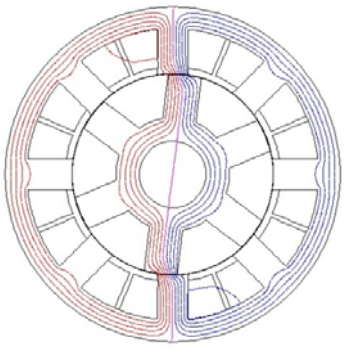

Figure 2. Flux lines for the rotational movement.

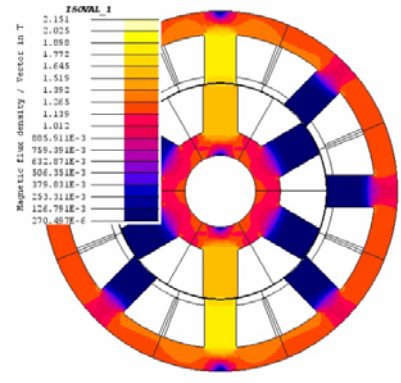

Figure 3. 2D flux density map.

The dynamic behaviour of the machine in study is simulated using co-simulation technology by coupling toghether Flux 2D and Matlab/Simulink by means of Flux-to-Simulink technique [6].

The simulation of the power converter was performed by using an electrical circuit built up in Electriflux (Flux 2D's circuit editor) attached to the FEM model of the machine. The electrical circuit corresponding to each phase is modeled using two electrical coils, corresponding to the "come and go" sides of the winding [7].

In the electric circuit model the power switches were replaced by resistors. Their resistances are set at a high value (OFF state of the switch) or at a low one, corresponding to the ON state of the transistor.

The main window of simulation program is given in Fig. 4.

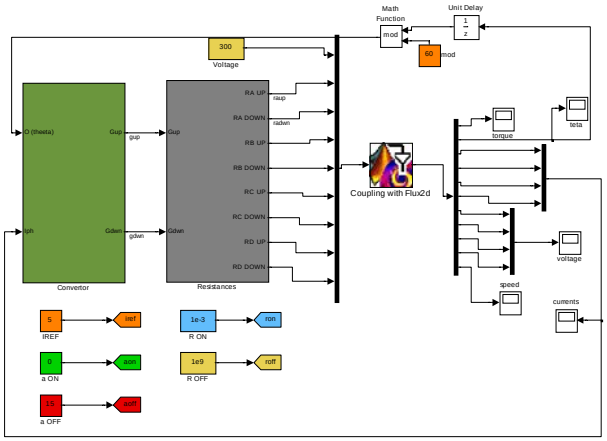

Figure 4. The main window of the simulation program.

As it can be seen the coupling program contains two main blocks. The green block generates the phase currents using the hysteresis current control technique and also (function of the rotor position) computes the firing angles. Based on the signals received from this block, the grey block is seting the values of the comand resistances for each phase.

When a step is computed in Simulink the results are sent to the Flux 2D model, and the response (concerning the position, torque, voltage and phase currents) is obtained after the field computations, and is sent back to the Simulink model.

The next step is computed based on these feedback values, so the system operates in closed reaction loop. The firing angles are defined upon the rotor position and the maximum phase current. The ON/OFF signals sent to the switches are: the resistance of 100 kΩ (OFF state) or 0.001 Ω (ON state).

The obtained results for the dynamic simulations (phase currents, voltage, torque and angular displacement) are given in Fig. 5 and it can be stated that the obtained torque value is quite the same as computed.

Also the torque and flux curves versus current and rotor position given in Figs. 6 and 7 were obtained by means of field computations.

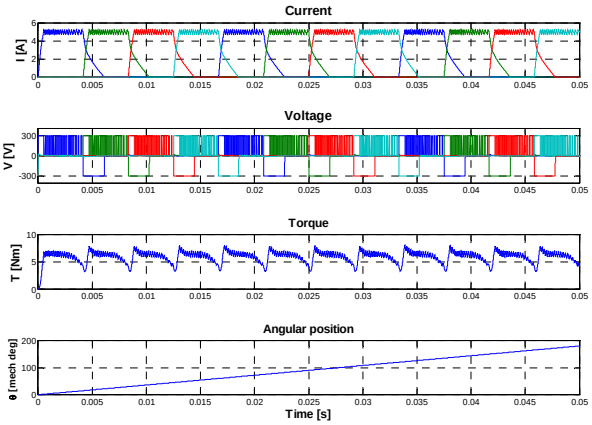

Figure 5. The obtained results for dynamic simulations.

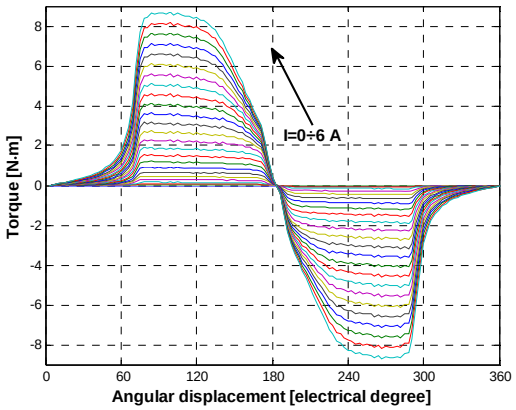

Figure 6. Torque vs. position characteristics.

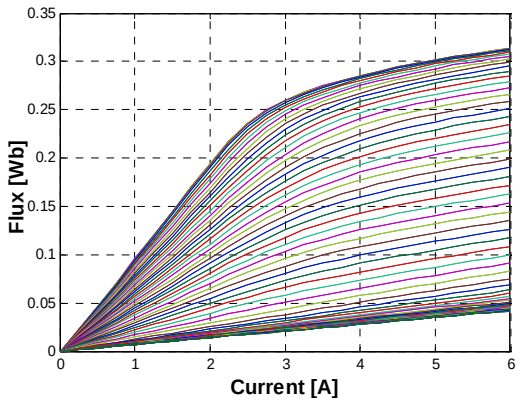

Figure 7. Flux vs. current characteristics.

These curves will be used to model the machine in MATLAB-Simulink. They will be implemented in the simulation program as look-up table blocks [8].

C. The linear motion

If linear movement is required the proposed motor will work similarly to a linear SRM. Upon the required direction of movement and the needed thrust one or more phases of a stator stack will be fed. The rotor stack will be aligned upon the variable reluctance principle with that stator's poles [9].

To simulate the translation motion, obligatory three dimensional field computations must be performed due to the complex 3D flux paths inside the motor. These computations were performed by using Flux 3D.

Due to the large amount of 3D mesh elements only a half of the motor was modelled and an adequate periodicity function was assumed.

The linear movement can be achieved using two feeding methods, depending on the needed tangential force. The first one is feeding the two coils that belong to the aligned poles and the second is to feed six of the stator coils at the same time [10].

For the first case the obtained results are given as colour maps of the magnetic flux density as shown in Fig. 8. It can be seen that the distribution of the flux density is quite closed to that computed via Flux 2D (Fig. 3).

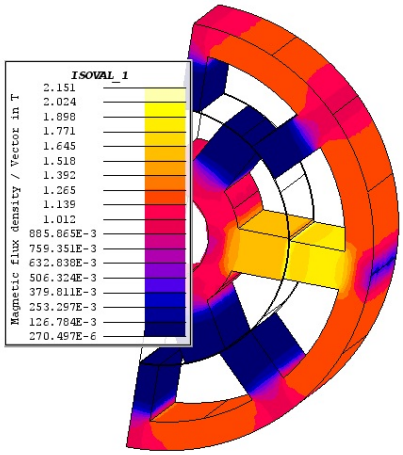

Figure 8. 3D colour map of flux density in aligned position.

The second way to feed the machine is to energize 6 of 8 coils of the stator at one time. As in the first case, two of the poles are aligned but four of them are in partially aligned position. It is considered that the torque produced by the four partially aligned poles will be equal but will have different orientation, hence the total torque will be null and no rotary movement will result.

The advantage of this type of feeding is obtaining a higher value for the tangential force. Thus, the way of feeding will be set according to the needed tangential force.

The flux density map for the second feeding method is given in Fig. 9.

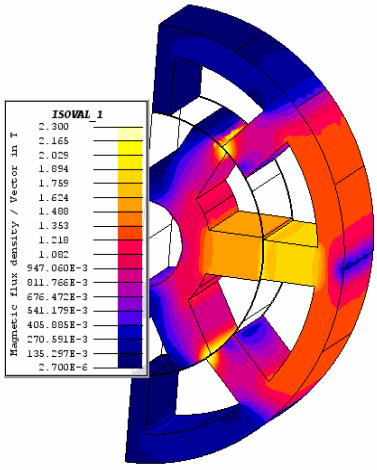

Figure 9. Flux density map for the second method of feeding.

As it can be seen that the flux path is different from the one seen in Fig. 8, which corresponds to the first feeding method. The flux is not circulating through all stator and rotor yoke. It has a shorted path created between the two partially aligned poles and the aligned pole.

It could be of interest to compare the magnetic flux densities obtained via 2D and 3D numeric field computations with those computed analytically during the design. The values obtained by means of the three methods are given in Table I.

Table I. Comparison of the magnetic flux densities

Region	3D analysis	2D analysis	Analytical
Stator pole	1.92 T	1.89 T	1.79 T
Stator yoke	1.33 T	1.29 T	1.27 T
Rotor pole	1.71 T	1.64 T	1.58 T
Rotor yoke	1.16 T	1.13 T	1.12 T
Air gap	1.83 T	1.81 T	1.75 T

As it can be seen the results obtained analytically and via the 2D numeric field computations are quite close. The 3D simulations are less precise because due to the long simulation times the mesh could not be enough refined.

Also the tangential force was determined for both feeding methods having the current set to 5A. A comparison between the obtained values of the tangential force is given in Fig. 10.

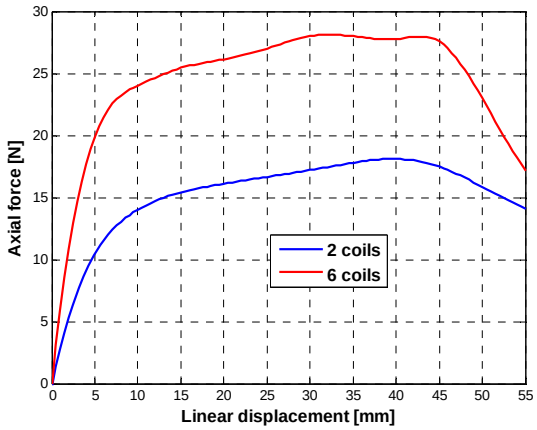

Figure 10. The tangential forces obtained by the two feeding methods.

As it can be seen, the value obtained for the second method is higher with 55% than the one obtained for the first method. Also it can be stated that for a higher tangential force the obtained graphical form is smoother.

Hence, the way of feeding the machine and the current value are set depending by the application where this machine is used and by needed values for torque and force.

III. CONCLUSIONS

The combination of a rotary and a linear movement on the same axis is frequently required in diverse automotive and industrial systems. For such applications the proposed rotary-linear SRM seems to be an excellent solution.

The 2D simulations compared with 3D simulation prove that the machine is well designed and the simulation results presented prove the usefulness of the proposed motor.

Also it is highlighted that different values of the tangential force can be achieved, function by the needs and application.

Future work is related to the development of a laboratory model of the machine and of the control system capable to enable well-coordinated dual motion.

ACKNOWLEDGEMENT

This paper was supported by the project "Doctoral studies in engineering sciences for developing the knowledge based society–SIDOC" contract no. POSDRU/88/1.5/S/60078, project co-funded from European Social Fund through Sectorial Operational Program Human Resources 2007-2013.

REFERENCES

[1] Z.Z. Liu, et al., "*Robust and precision motion control system of linear motor direct drive for high-speed XY table positioning mechanism*," IEEE Transactions on Industrial Electronics, vol. 52,2005, pp. 1357-1363.

[2] I. Boldea and S. A. Nasar, "*Linear Electric Actuators and Generators*," Cambridge University Press, 1997.

[3] J.F Pan, N.C. Cheung and G. Cao, "*Investigation of a rotary-linear switched reluctance motor*," in Proc. of the XIX International Conference on Electrical Machines (ICEM '2010), Rome (Italy), 2010, pp. 1-4.

[4] P. Bolognesi, O. Bruno, F. Papini, V. Biagini and L. Taponecco, "*A low-complexity rotary-linear motor useable for actuation of active wheels*," in Proc. of the International Symposium on Power Electronics Electrical Drives Automation and Motion (SPEEDAM '2010), Pisa (Italy), 2010, pp. 331-338.

[5] I.A. Viorel, et al., "*Speed-thrust control of a double sided linear switched reluctance motor (DSL-SRM)*," in Proc. of the 18th International Conference on Electrical Machines (ICEM '2008), Vilamoura (Portugal), 2008.

[6] M. Busi and S. Cadeau-Belliard, "*Induction Motor Drive using FLUX to Simulink Technology*," FLUX Magazine, no. 47 (January 2005), pp. 15-17.

[7] M. Ruba, I. Benţia. and L.Szabó, "*Novel Modular Fault Tolerant Switched Reluctance Machine for Reliable Factory Automation Systems*," Proc. of the 2010 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR '2010) THETA 17, Cluj (Romania), 2010, Tome III, pp. 47-52.

[8] L.Szabó, I. Benţia, and M. Ruba, "*Dual motion switched reluctance motor for advanced industrial applications*," Proc. of OPTIM '2012 (in press).

[9] I. Benţia, L.Szabó and M. Ruba, "*On the control of a rotary-linear switched reluctance motor*," in Proc. of the 5th International Symposium on Computational Intelligence and Intelligent Informatics (ISCIII '2011), Floriana (Malta), 2011, pp. 41-46.

[10] I. Benţia, L.Szabó and M. Ruba, "*On a rotary-linear switched reluctance motor*," Proc. of the SPEEDAM '2012 Conference (in press).

Real-time Simulation Environment for Embryonic Networks

CHINDRIŞ Virgil, SZÁSZ Csaba

Techincal University of Cluj-Napoca, Romania,
Department of Electrical Machines and Drives, Faculty of Electrical Engineering,
Memorandumului nr. 28, RO-400114 Cluj-Napoca, Romania, E-Mail: virgil.chindris@mae.utcluj.ro

***Abstract* – Throughout the time, evolution processes of biological organisms are characterized by self-healing, adaptation and not the last, surviving abilities. Mimicking all these features in VLSI circuits, this has lead to technological improvements that make possible the designing and building of fault-tolerant systems. Because complexity is almost a requirement of these systems, in order to exhibit the above mentioned features, designing and simulating them efficiently, require a tailored modeling and simulation environment. By studying the design and implementation flow of embryonic systems, it can be concluded that a modeling and simulation environment has to have the capabilities of modeling through a programming language, as close as possible to the implementation methods, and in the same time the power of simulating in real-time. The paper focuses on presenting a new modeling and simulation environment, giving an example of a structure for an embryonic system.**

Keywords: ***Real-time; Simulation environment; Embryonic system; Fault-tolerant system; Bio-inspired***

I. INTRODUCTION

All the living organisms are endowed by nature with distinct evolution capabilities like self-healing abilities, surviving and adaptation, encoding them genetically. To design complex fault-tolerant hardware, these capabilities are a key feature to include [1, 2]. Most of the computer aided design and simulation environments, typically involved in these processes, either natural or replicated in hardware, approach them in a classical fashion, where the simulation runs for a while and after done, the results are presented for analysis or further post processing. The design and build time of a bio-inspired hardware system is drastically reduced if the simulation runs in real-time, presenting to the user how a process evolves and where or when the process fails [3]. In such a system, a bug in software code is very difficult if not impossible to spot or trace when using a non real-time simulation. In a classical simulation, although the results present various design defects, the complexity of the system hinders the defect tracing.

A major advantage of a real-time simulation is that it can accept external stimuli or even altering parts of the simulated model, during run time [3]. Such a feature is beneficial either for design improvements or debugging. From different points of view, the simulation of embryonic systems raise several types of difficulties. One is regarding the incapability of most present software tools available on the market to simulate multiple digital structures with complex VLSI processors, organized in network architectures. Although there are some tools capable of simulating networks in real-time, their internal models are described through electronic schematics, which make them difficult to develop compared to a coding approach. Beside of this problem, also those simulation environments, which are suited for digital circuits synthesizing, usually compute (through numerical integration algorithms) only analogical magnitudes like currents and voltages, and such processes increase significantly the simulation time, excluding them from the real-time category. For this reason, new research instruments and advanced study tools are far welcome for engineers involved in complex digital systems simulation and development.

The tradeoff between different combination of features constrains the user somehow when he or she must take a decision when choosing the best suited simulation environment.

It is true that fault-tolerant hardware systems include parts that can't be simulated in real-time or simply has no relevance for this type of simulation, but practice shows that in most cases these parts can have a simplified model or even excluded from the real-time simulation. The real-time modeling and simulation environment, named *DigChipSim*, presented in this paper is aimed to overcome some of the above mentioned shortcomings. Also, it is given an example of how a fault-tolerant system is simulated using the *DigChipSim* environment.

II. THE *DIGCHIPSIM* ENVIRONMENT

The most common units of a simulation environment are the design, simulation and post processing. The *DigChipSim* environment presents a versatility in design, characterized by allowing the user to describe the models in a programming language, easing their development, use custom GUIs or even write external simulators in their preferred design environment [4].

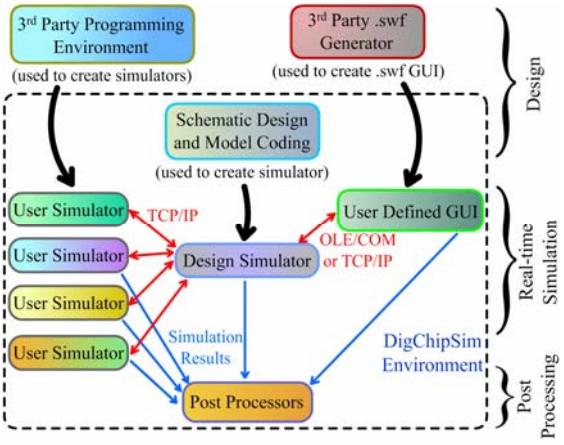

Figure 1. Structure of the *DigChipSim* environment.

The structure of the simulation environment of *DigChipSim*, written in *Turbo Delphi* [5], presented in Fig. 1, shows the three mentioned units: design, simulation and post processing. The Design unit from Fig.1 includes the environment's design application, named *Digital Chip Simulator*, used for schematic design and model coding and two design applications or environments which are not actually parts of the *DigChipSim* environment, but are required for advanced use. These last two design environments are 3rd party environments, used for writing custom GUIs and custom simulators. The thick arrows from Fig. 1 show the design application or environment [5, 6] that is used for the different simulators of the *DigChipSim* environment.

In the real-time simulation unit, there are included a design simulator, which is the main simulator, several user simulators and a custom GUI simulator [6]. The input/output interface from the design simulator is limited because complex GUI interfaces or simulators built by the user are always more suited for the simulated design than any other solution. The simulator that loads custom user GUIs is called *FlashWin* and loads GUIs in a *swf* (*Shockwave Flash*) format [7].

The last unit, called post processing, comprise several applications that process the simulation results, data that is not needed to be processed in real-time during simulation.

The main idea of the simulated application, given as an example, is based on an artificial organism with a general structure like in Fig. 2.

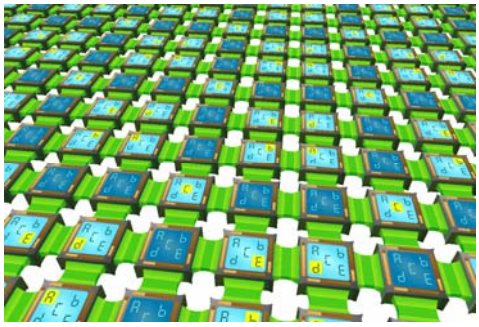

Figure 2. Structure of an artificial organism [9].

An artificial organism, like that presented in Fig. 2, is comprised from a large number of artificial cells, organized as active cells and spare cells [8]. For the presented example, the application uses only a cluster of cells, made by a 3x3 matrix and several external devices, resulting in an embryonic machine. The structure of this embryonic machine is presented in Fig. 3.

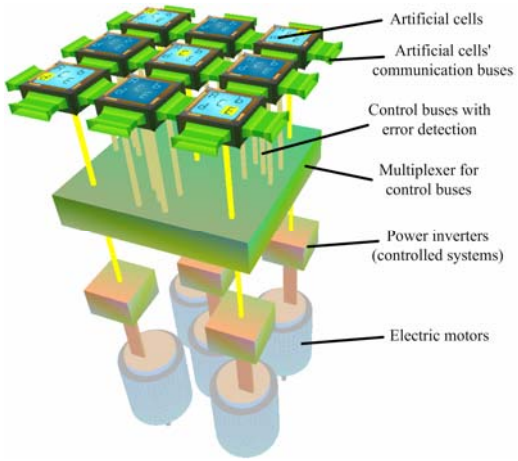

Figure 3. The embryonic machine used for simulation.

For the design/modeling step of the process, the main window of the *Digital Chip Simulator* application from the *DigChipSim* environment is presented in Fig. 4.

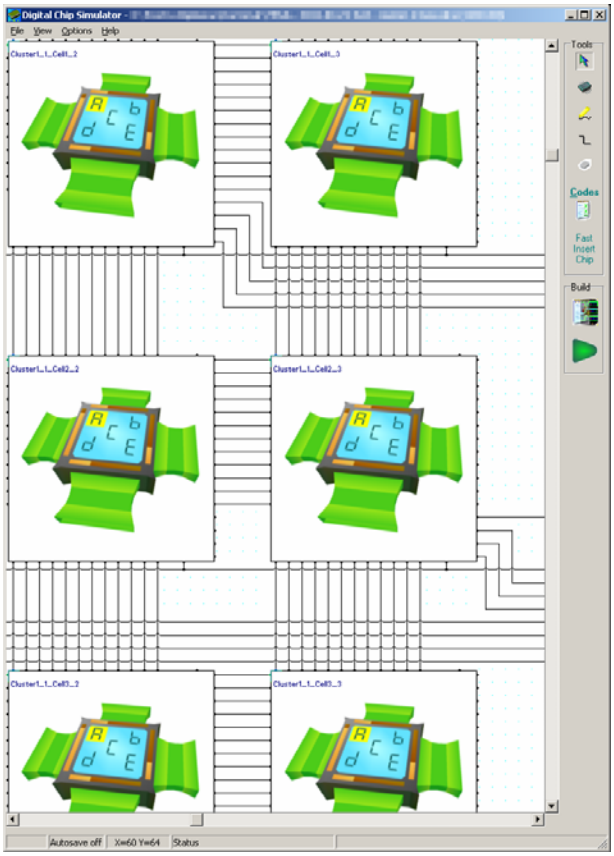

Figure 4. The main window of the *Digital Chip Simulator* application.

Fig. 4 shows the main window of *Digital Chip Simulator* in which there can be seen a design board, a toolbar at the right hand and the program's window menu. The first stage of a design in *DigChipSim* is capturing the circuit, which will be simulated later. For simplicity, a circuit designed in *DigChipSim*, consists only of integrated circuits, no matter their complexity and connection wires between them. The integrated circuits used in a design will be called further chips or blocks. To have a fully customizable design, the user is able to model the blocks by writing Pascal code for each block. After finishing circuit capture and blocks' modeling, the user hits the "build" button to make the design available for simulation.

Using the chip editor presented in figure 5, there can be designed chips at a maximum size of 400 pins (100 pins on each side). Based on a list of variables, took from the chip model, the user assigns each pin to a variable, either in a manual fashion or using the *AutoArrange* feature. The algorithm behind this feature takes the list of variables which will be assigned to pins, and based on their direction (either input or output) the input variables are assigned to the pins from left and top sides and the output variables are assigned to right and bottom sides of the chip. Thus, the chip is sized on x and y dimensions accordingly.

Using the chip editor, there can be also loaded pictures that will be overlayed on the chips on the design board. This feature helps the user to easily understand the role of each chip in a big schematic. The overlayed images are automatically sized to the owning chip.

In the same editing window, the user assigns a chip model for the chip being edited [10].

The chip editor allows also simple transformations of the chip, like clockwise / counter-clockwise rotations, horizontal and vertical rotations and X / Y sizing.

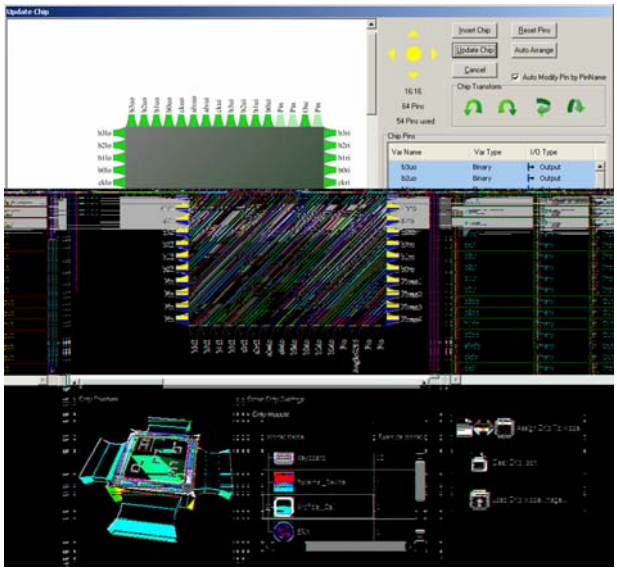

Figure 5. The chip editor of the *Digital Chip Simulator* application.

Editing the chip models is done using the *Chip Models and Code Editor* window from the *Digital Chip Simulator* application, presented in Fig. 6.

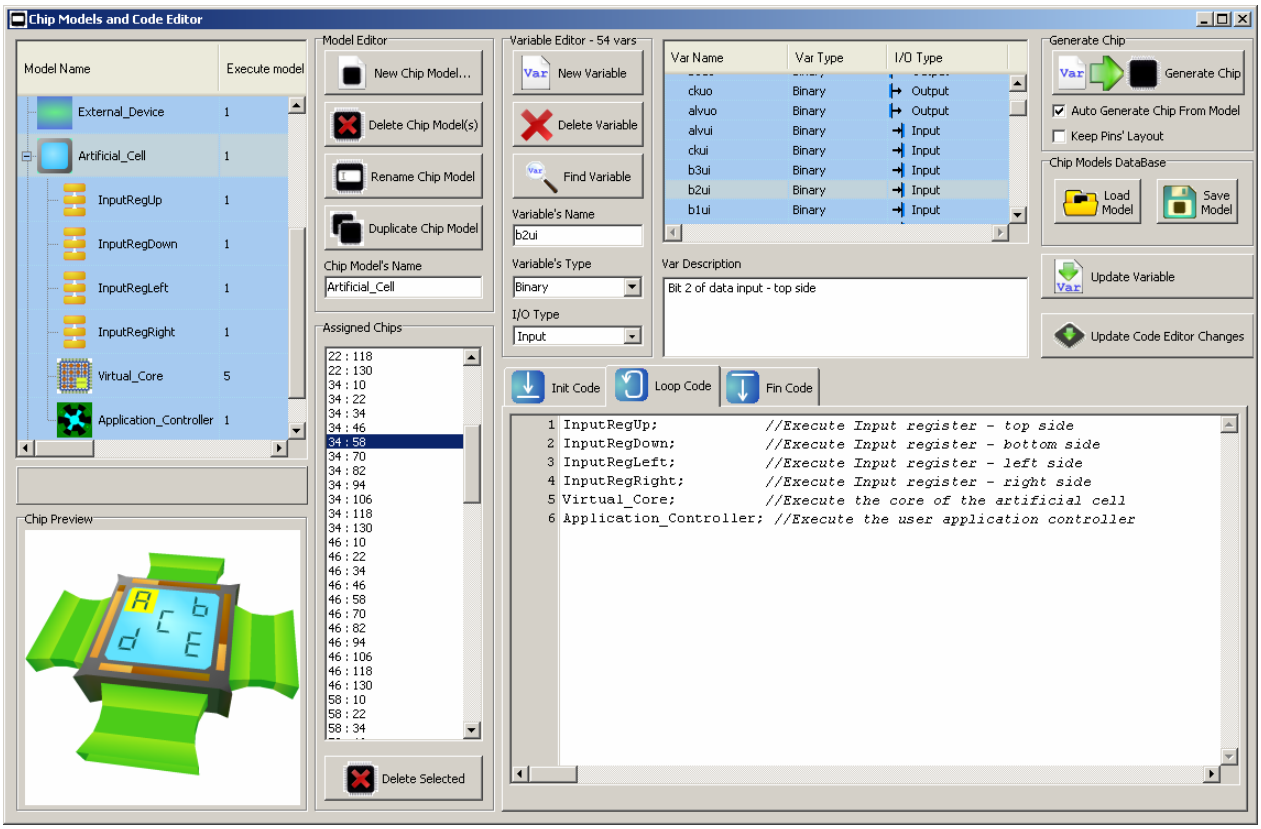

Figure 6. The *Chip Models and Code Editor* window of the *Digital Chip Simulator* application.

Comparing to most of the real-time modeling and simulation software available on the market, one of the strengths of *DigChipSim* environment is that it allows the user to describe the models using programming code [11]. The models are organized in a tree fashion, with the top-level models being assigned to chips on the design board and the other models working as subprograms/subcircuits.

Every model may have code for initialization, for loop and for finalization. The code describing a subcircuit can be called as a subprogram from its upper-level parent model. The loop code from the top-level models is executed in a sequential fashion, directly controlled by the simulator.

A feature required by the simulation of an embryonic system is that the environment should allow for assigning multiple chips to the same model, for example, giving to all the artificial cells from the Fig. 2, the same code. Each chip model has its own code and variables which are dynamically created for each assigned chip. The variables of input and output type must be assigned to the chip's pins in order to be used as communication lines. Variables that are of internal type can't be assigned to the chip's pins and are used for internal purposes only.

Another advanced feature is the execution rate of each model when there is needed to simulate various parts of the model at different speed. For example, in the artificial cells used, there are subcircuits that have to run at full speed and a core processor which runs slower. This translates in a physical counterpart that behaves like a combinatorial logic from a shift register, for example, and a sequential logic from a virtual processor.

Different other useful features of the chip model and code editor include previewing the overlay images of assigned chips, saving / loading a chip model to / from file, adding comments to variables and calling the chip editor to automatically generate the chip based on the variables' list.

III. REAL-TIME SIMULATION UNITS

The simulation principle of the presented simulator consists of a data transfer method that synchronizes data at chip transfer level by grouping simulation tasks in two main categories: chip models execution and connection execution. This approach requires that all inputs of the simulated models remain constant during models' execution and all outputs remain constant during connections' execution. Thus, no matter the complexity of a model in regard to other model, the execution time of each model lasts exactly one iteration, which means a simulation step.

The execution of a model consists of computing all its state equations by taking a set of inputs and updating its outputs accordingly. The graphical updates and communication with other simulators are also done here. Executing the connections (that link the models) means updating models' inputs with outputs from other models. Fig. 7 shows a flowchart of the simulation principle presented in this paper.

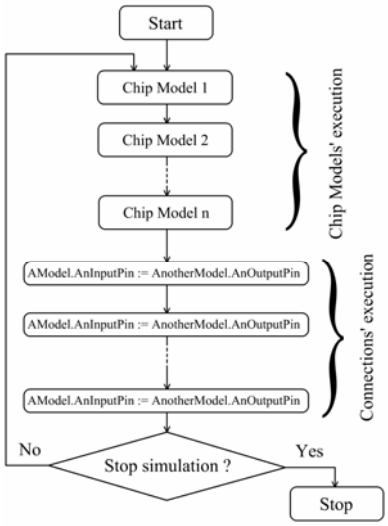

Figure 7. The simulation flow in *DigChipSim.*

In contrast with other simulators, this one does not mix the models' and the connections' executions thus making it capable of modeling digital circuits of a wide range of complexities. By executing the models on multiple processors and synchronizing their results after execution, there can be achieved a high simulation performance.

This simulator can also be used in analog models that do not require simulation of currents but only voltages, however, the support for analog simulations is limited. The main disadvantage of this simulation principle is that it can't be used for models where the precise timing is a critical factor. For those types of simulations, the best suited are the non real-time ones.

For the reason of being able to let the user interract in a friendly manner with the simulation, the *DigChipSim* environment provides the *FlashWin* simulator, which is capable of loading user defined GUIs, designed in a 3rd party environment. For example, a user defined graphical interface that can be loaded in *FlashWin* is a *.swf (ShockWave Flash)* file which is based on named objects. In this way, the user can design an application specific GUI that is the closest model to its real counterpart as shown in Fig. 8.

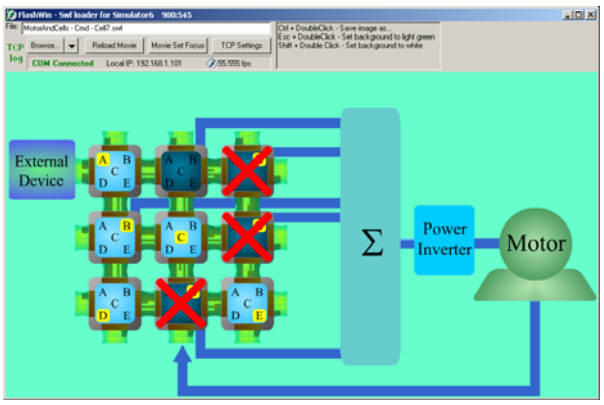

Figure 8. The *FlashWin* application from the *DigChipSim* environment.

This interface accepts simulation stimuli from the user either by using the mouse or using the keyboard. The simulation interface loaded in *FlashWin*, presented in Fig. 8, lets the user "kill" the artificial cells with a mouse click. The feedback to the user becomes complete by graphically updating the state of the simulated models. For example, the control wires glow if their state is active; the artificial cells show the active genes.

Another advantage of this approach is the easiness in debugging because the user can see in real-time, how every part of the simulated model works. The *FlashWin* simulator can be run on a different computer because it can be connected to the main simulator of *DigChipSim* through a TCP/IP link [12]. For local connections, *FlashWin* can use a COM (*Component Object Model*) interface [13], which is more accessible than TCP/IP.

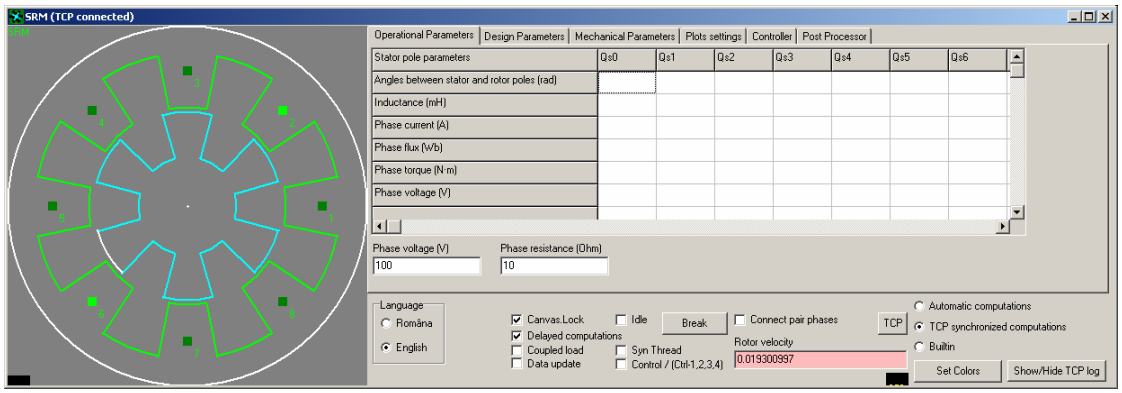

Figure 9. A custom user real-time simulator designed for simulating a switched reluctance machine [14].

For the embryonic machine presented further in this paper, a custom user defined simulator was used, given in Fig. 9. This is a standalone program like those presented above, which is designed to simulate a switched reluctance machine in real-time for low velocities [14]. This kind of simulator can be tailored even more than the user defined GUI loaded in *FlashWin*, because it is built from scratch in a programming environment. The user defined simulator communicates with the main simulator of *DigChipSim* through a TCP/IP connection, making it able to run on a different computer [15].

IV. A FAULT-TOLERANT EMBRYONIC SYSTEM SIMULATED IN *DIGCHIPSIM*

The artificial cells, briefly presented in Fig. 2 and Fig. 3, are used as motor controllers for a 4-phased, 8/6 switched reluctance machine. From the block diagram given in Fig. 8, loaded as a user defined GUI in *FlashWin*, it can be seen that there is a 3x3 matrix of artificial cells, constituting a cluster, an external device, used as interface to a personal computer, a multiplexer, named Σ, for control buses, a power inverter and a motor. The cell from the top-right corner, having the matrix notation Cell_1_3, is assigned as the active cell to drive the motor. Its control algorithm, also called *artificial gene* is highlighted with letter "B". Other active cells highlight other genes, for their own control algorithms. The total number of genes available in a cluster is given by the number of active cells, which is 5. The other 4 artificial cells are used as spare cells and each of these spare cells can replace any other cell by simply activating the appropriate gene.

Fig. 10 presents a simulation result in which the cell 1_3, used to drive the motor, is faulted at some point and it is replaced by spare the cell 2_3. [8]

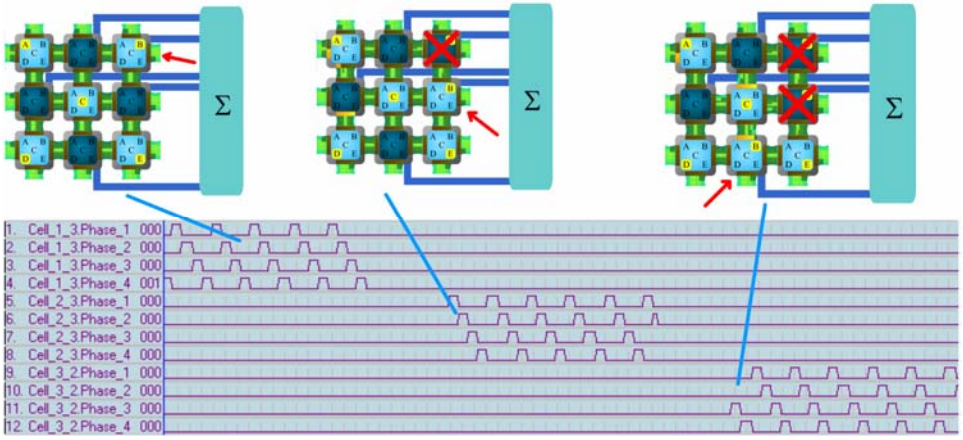

Figure 10. Motor phases waveforms in different stages of the artificial cell replacing process.

The replacing process consists of detecting the fault by monitoring the control bus of the active cell by the spare cell and activating the appropriate gene to continue the control. Later, also the spare cell 2_3 is faulted and it becomes replaced by the spare cell 3_2. Although Fig. 10 does not present any other cell faults inside the cluster, they can be handled further because there are still two remaining spare cells. The red arrows from Fig. 10 show which cell is actively driving the control bus for the motor phases at a different stage of the simulation.

Another important information given by Fig. 10 is represented by the motor phases waveforms, which show that in case of a cell fault, the spare cell generates as quick as possible the control signals to keep the motor running. In any case, it can be observed that the embryonic array, represented by the artificial cells, keeps its immunity during this process, showing active the same genotype (*A, B, C, D*, and *E*).

Most of the modeling and simulation details of the presented example, are given in figures Fig. 3 to Fig. 9, took as snapshots after the modeling stage. There are shown details from electronic schematics to model's architecture and even simulation flow.

In the schematic from Fig. 8, the "Power Inverter" is a component that appears in a physical application but can be omitted in a simplified model, where the motor phases can be controlled directly with logic signals, e.g.: ON/OFF. Because the simulated switched reluctance motor (Fig. 9) can be controlled with such signals, the "Power Inverter" may have no model. In the presented model, the phase currents are not used because the motor is controlled only from its rotor's position. This feedback must go to the artificial cell that is responsible for controlling the motor (according to its active gene) and all the spare cells. In case of a fault in the artificial cell network, there are spare cells ready to replace the faulted cell and continue the controlling process by being connected both to the electrical machine and the sensors.

V. CONCLUSIONS

The paper is focused on presenting a real-time modeling and simulation environment specially tailored for embryonic systems, named *DigChipSim*. This environment can be used for modeling by designing electronic schematics that contain models described in a programming language. The simulation is done in a real-time fashion and allows also processing of the results.

The presented theoretical approaches and real-time simulation software toolkit could be a useful support for computer-aided modeling and simulation of complex bio-inspired hardware systems or sophisticated embryonic processes. Implementing basic properties of living organisms like self-healing or surviving in hardware architectures, highly reliable embryonic systems become possible to be developed.

ACKNOWLEDGMENTS

This paper was supported by the project "Doctoral studies in engineering sciences for developing the knowledge based society-SIDOC" contract no. POSDRU/88/1.5/S/60078, project co-funded from European Social Fund through Sectorial Operational Program Human Resources 2007-2013.

REFERENCES

[1] Caponetto, R.; G. Dongola; and L. Fortuna; "A New Class of Fault-Tolerant Systems: FPGA Implementation of Bio-Inspired Self-Reparing System" Proceedings of the 15th Mediterranean Conference on Control & Automation, Athens, Greece, TO1-028. 2007

[2] Mange, D.; A. Stauffer; E. Petraglio; G. Tempesti; "Artificial cell division" Biosystems, Vol. 76, no 1-3, 157 167. 2004

[3] Bauer, P. and P.J. van Duijsen. "Challenges and Advances in Simulation," Proceedings of the 36th Annual IEEE Power Electronic Specialists Conference (PESC '05), Recife, (Brazil), pp. 1030-1036. 2005

[4] Cantu, M. Mastering Borland Delphi 2005. SYBEX, Indianapolis (USA). 2005

[5] Borland Inc. Developer Studio 2006 Reference. Delphi Language Guide, C++ Language Guide, Together Reference. Borland Software Corporation, Scotts Valley (USA). 2006

[6] Adobe Flash, http://www.adobe.com/products/flash.html

[7] "SWF File Format Specification Version 10", http://www.adobe.com/content/dam/Adobe/en/devnet/swf/pdf/swf_file_format_spec_v10.pdf

[8] Szász, Cs.; V. Chindriş; 2010. "Development of Hardware Redundant Embryonic Structure for High Reliability Control Applications" 12th International Conference on Optimization of Electrical and Electronic Equipment, OPTIM 2010 Brasov, Romania, ISSN: 1842-0133, ISBN: 978-973-131-080-0, IEEE 978-1-4244-7020-4, 728-733. 2010

[9] Chindriş V., Szász Cs. – Artificial Genes Implementation Upon FPGA-Based Embryonic Network, 25th European Conference on Modeling and Simulation, ECMS, Krakow, Poland, ISBN: 978-0-9564944-2-9, pp. 153-158, 2011.

[10] Garrido, J.M. Object Oriented Simulation. A Modeling and Programming Perspective. Springer, Dordrecht (Germany). 2009

[11] Bauer, P. and P.J. van Duijsen. "Challenges and Advances in Simulation," Proceedings of the 36th Annual IEEE Power Electronic Specialists Conference (PESC '05), Recife, (Brazil), pp. 1030-1036, 2005

[12] Kozierok, C.M. The TCP/IP guide: a comprehensive, illustrated Internet protocols reference. No Starch Press, San Francisco (USA). 2005

[13] "Component Object Model, Microsoft COM Technologies", http://www.microsoft.com/com/

[14] Chindris, V., Terec, R., Ruba, M., Szabó, L., Rafajdus, P.: Useful Software Tool for Simulating Switched Reluctance Motors. In: Proceedings of the 25th European Conference on Modelling and Simulation (ECMS '2011) Krakow (Poland), pp. 216-221. 2011

[15] Szabó, L., Ruba, M.: Using co-simulations in fault tolerant machine's study. In: 23rd European Conference on Modelling and Simulation, Madrid, pp. 756 762. 2009

A Control Design for Grinding Systems with Feedforward Compensation

COSTEA Claudiu Raul
Department of the Engineering of the Automated Systems and Management,
University of Oradea , Faculty of Electrical Engineering and Information Technology,
Str. Universităţii 1, 410087, Oradea, Romania,
E-mail: ccostea@uoradea.ro

***Abstract* – The object of the present study is to design a control system for milling systems. An analyze of the dynamic behavior is described, aimed at designing a control system. The control system that is presented uses PID control with feedforward compensation.**

***Keywords:* ball mill; cement mill; feedforward; mill product; grinding circuit.**

I. INTRODUCTION

"Early hydraulic cements, such as those of James Parker, James Frost and Joseph Aspdin were relatively soft and readily ground by the primitive technology of the day, using flat millstones" [1]. "The emergence of Portland cement in the 1840s made grinding considerably more difficult, because the clinker produced by the kiln is often as hard as the millstone material" [1]. "Because of this, cement continued to be ground very coarsely (typically 20% over 100 μm particle diameter) until better grinding technology became available" [1]. "Besides producing un-reactive cement with slow strength growth, this exacerbated the problem of unsoundness" [1].

"This disruptive expansion is caused by hydration of large particles of calcium oxide" [1]. "Fine grinding lessens this effect, and early cements had to be stored for several months to give the calcium oxide time to hydrate before it was fit for sale" [1]. "From 1885 onward, the development of specialized steel led to the development of new forms of grinding equipment, and from this point onward, the typical fineness of cement began a steady rise" [1].

"The progressive reduction in the proportion of larger, un-reactive cement particles has been partially responsible for the fourfold increase in the strength of Portland cement during the twentieth century" [1]. "The recent history of the technology has been mainly concerned with reducing the energy consumption of the grinding process" [1].

The cement grinding aids are the additional materials, admixed in small amounts during the cement grinding process, which can significantly improve the grinding efficiency, reduce energy consumption, without compromising the performance of the cement.

II. PURPOSE OF A BALL MILL IN CEMENT MANUFACTURING PROCESS

"Ball mills rotate around a horizontal axis, partially filled with the material to be ground plus the grinding medium" [2]. "Different materials are used as media, including ceramic balls, flint pebbles and stainless steel balls" [2]. "An internal cascading effect reduces the material to a fine powder. Industrial ball mills can operate continuously, fed at one end and discharged at the other end" [2].

"Ball mills are also used in pyrotechnics and the manufacture of black powder , but cannot be used in the preparation of some pyrotechnic mixtures such as flash powder because of their sensitivity to impact" [2]. High-quality ball mills are potentially expensive and can grind mixture particles, enormously increasing surface area and reaction rates, the grinding works on principle of critical speed and the critical speed can be understood as that speed after which the steel balls start rotating along the direction of the cylindrical device; thus causing no further grinding [2].

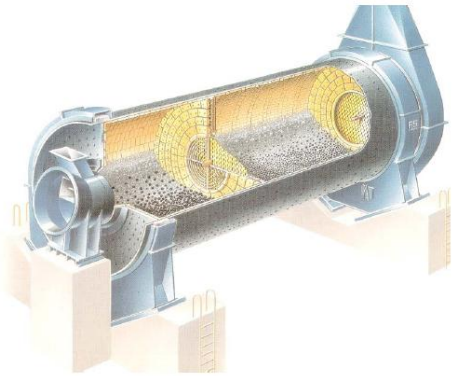

Figure 1. The ball mill.

The ball mill (fig.1) is still the most accepted element in the cement grinding. The reasons are: high reliability, simple operation, easy to maintain. "Ball mill is an efficient tool for grinding many materials into fine powder. It is used to grind many kinds of mine and other materials, or to select the mine" [2]. It is used in building material, chemical industry, etc.

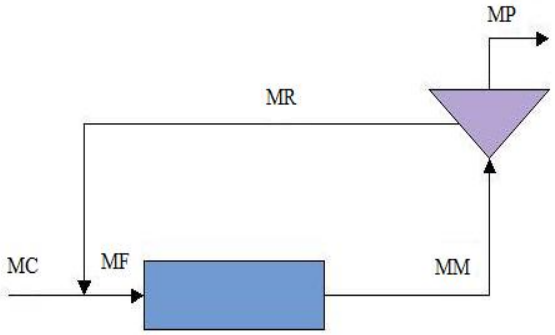

Figure 2. Industrial grinding circuit.

Fig. 2 illustrates a ball mill in closed loop with an air separator. The feed flow MF made up of clinker and other raw material components, enters the rotating mill where it is ground by steel balls. The mill product MM is transported into the separator where it is subdivised into a flow of fine particles, the final product denoted by MP, and a flow of rejected particles MR, which is recirculated to the mill inlet. The sum of the raw material flow MC and the recirculated flow MR constitues the total feed to the mill MF [3].

$$M_C + M_R = M_F \tag{1}$$

where:

M_C represent raw material flow ;

M_R represent a flow of rejected particles, which is recirculated to the mill inlet ;

M_F represent feed flow ;

Expression (1) illustrates the elements from the mill inlet. Expression (2) illustrates the elements form the separator.

$$M_M = M_R + M_P \tag{2}$$

where:

M_M represent the mill product ;

M_P represent final product ;

M_R represent a flow of rejected particles, which is recirculated to the mill inlet ;

III. THE SYSTEM WITH FEEDFORWARD COMPENSATION

"Feedforward control is generally agreed to be the single most useful concept in practical control system design beyond the use of elementary feedback ideas". [9]. Clearly if one can measure up-stream disturbances, then by feeding these forward, one can take anticipatory control action which pre-empts the disturbance affecting the process.

The feedforward compensation is used to provide the fastest possible response to dynamic changes in the input signal.

The basic principle of this scheme is the compensation of the various of the feed flow rate induced by the recirculated flow rate [3]:

$$\Delta M_C = -\alpha \cdot \Delta M_R \tag{3}$$

with $0 < \alpha < 1$.

"As a feedforward compensation does not ensure steady-state error, a control loop is implemented, which uses the recirculated flow rate and the mill flow rate to reconstruct the feed flow rate" [3].

One of the simplest and advanced control strategies is the PID control with feedforward compensation [11]. The PID (Proportion Integral Differential) controller is described by [11]:

$$u(t) = k_p \cdot e(t) + k_d \cdot \frac{de(t)}{dt} + k_i \int e(t)dt \tag{4}$$

where k_p, k_i, k_d are the proportional, derivative and integral gains.

The PID controller gains were tuned using the Ziegler–Nichols step response method, which correlates the controller parameters to features of the step response, with additional manual fine-tuning [4].

„The PID controller is designed to provide all of the functions required to implement a complete high performance, closed loop"[12].

The PID controller it is a simple strucutre, stable, reliable, easy to adjust to become one of the main techniques of industrial control [5]. One of the PID parameter adjustment and feedforward compensation method to control the system, can achieve more satisfactory control effect. [5]

The PID-controller design in Simulink it is show in Fig.3:

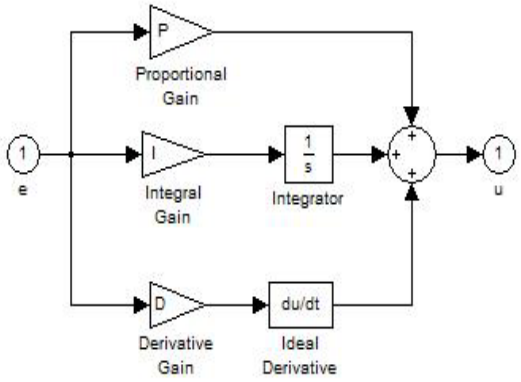

Figure 3. PID-controller.

The PID controller calculation involves three separate parameters: "the proportional, the integral and derivative values: the *proportional* value determines the reaction to the current error, the *integral* value determines the reaction based on the sum of recent errors, and the *derivative* value determines the reaction based on the rate at which the error has been changing" [19].

IV. SIMULATION RESULTS

The Simulink model based on (1), (2) and (3) is shown in the Fig. 4.

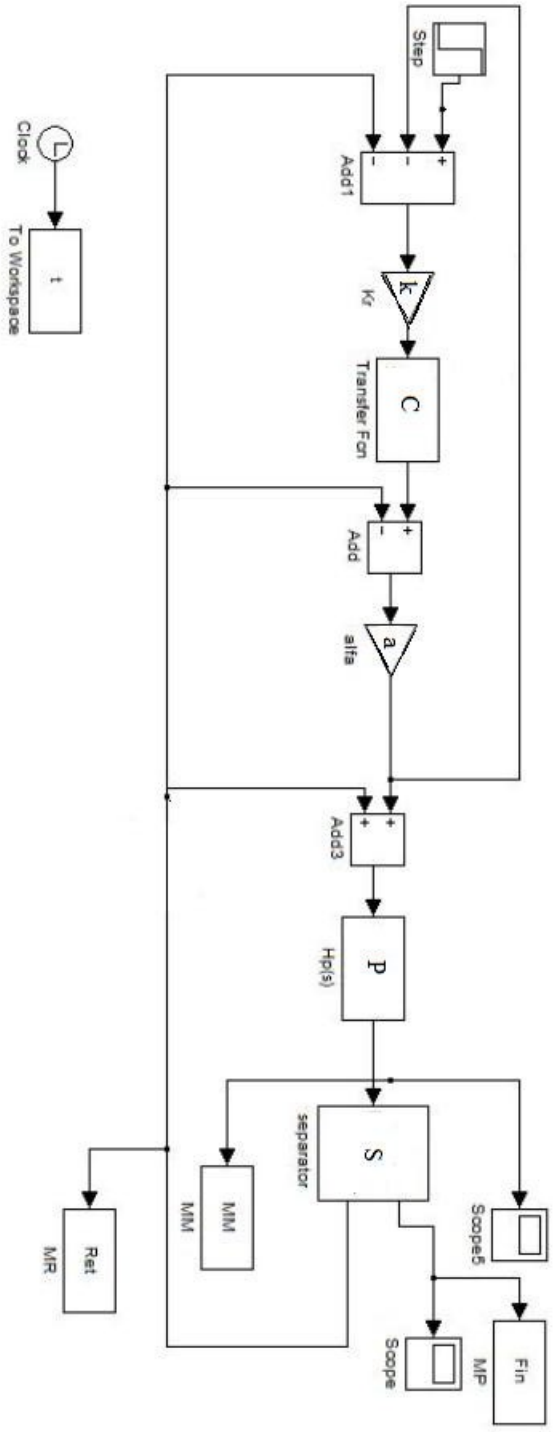

Figure 4. Simulink model.

This Simulink model containes a few block from varoius libraries, such as: sources blocks, for this case it use a Step block who generate step function, Scope who display signals generated during simulation, Add, Subsystems.

In Fig. 4, it is used the following symbols:

- k is the gain,
- C is the controller,
- a is the compensation coefficient,
- P represent grinding process,
- S represent separation process,
- Step block gives the ideal value for the system output, namely the reference value for mill product,
- MM is the mill product,
- Fin is the final product,
- Ret is the flow of rejected particles, which is recirculated to the mill inlet,
- t is the time;

"MATLAB is a numerical computing environment and programming language" [16]. "Maintained by The MathWorks,MATLAB allows easy matrix manipulation, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs in other languages" [16]. "Although it is numeric only, an optional toolbox uses the MuPAD symbolic engine, allowing access to computer algebra capabilities" [16]. "An additional package, Simulink, adds graphical multidomain simulation and Model-Based Design for dynamic and embedded systems" [16].

"Simulink is developed by The MathWorks, it is a commercial tool for modeling, simulating and analyzing multidomain dynamic systems" [17]. "It is primary interface is a graphical block diagramming tool and a customizable set of block libraries" [17]. "It offers tight integration with the rest of the MATLAB environment and can either drive MATLAB or be scripted from it. Simulink is widely used in control theory and digital signal processing for multidomain simulation and design" [17].

The dynamic behavior of a cement mill is simulated using MATLAB-SIMULINK. In the figures legend, have noted mill product with *MM*, final product (cement flow) with *MP* and recirculated flow with *MR*. Fig. 5 show variation of final product.

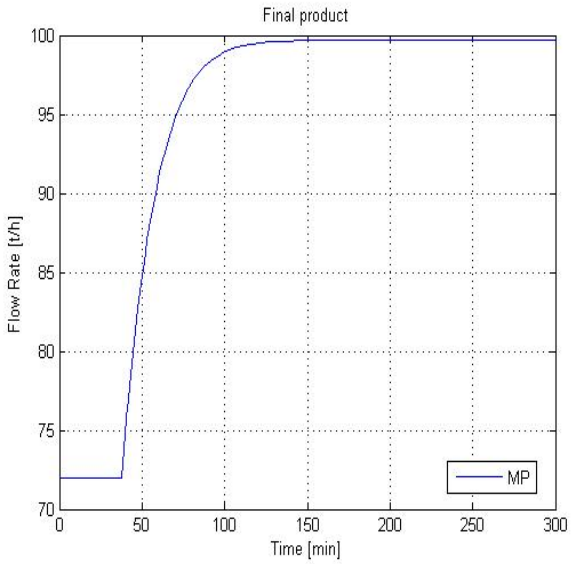

Figure 5. Final product.

Figures 6 and 7 show variation of the mill product, of the recirculation flow and of the final product in relation with time.

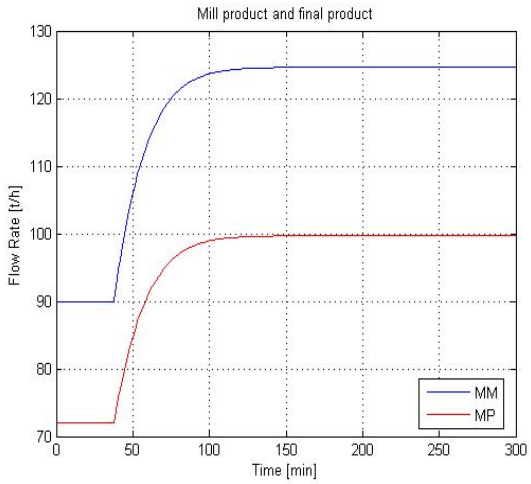

Figure 6. Mill product and cement flow.

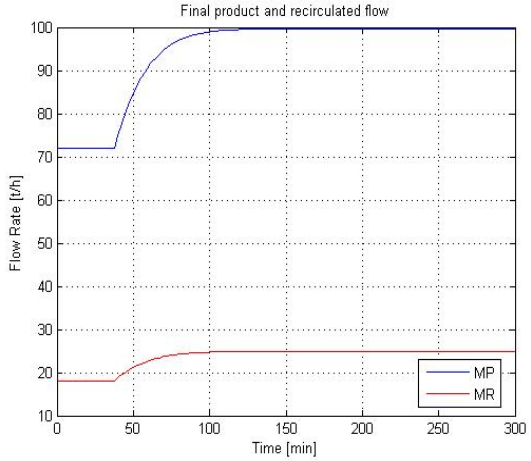

Figure 7. Final product and recirculated flow.

From Figures 5,6 and 7 can see the next things:

- mill product increase from 90 t/h to 125 t/h, after the step is applied ;
- cement product increase from 72 t/h to 100 t/h, after the step is applied ;
- recirculated flow increase from 18 t/h to 25 t/h, after the step is applied ;

V. CONCLUSIONS

It is presented a feedforward control of milling operations in order to improve productivity. The control system presented in this paper was designed and tested in simulation. Thanks to the process modeling and simulation other control strategies could also deeply investigated to a high energy consuming process as the grinding is.

REFERENCES

[1] http://en.wikipedia.org/wiki/Cement_mill
[2] http://en.wikipedia.org/wiki/Ball_mill
[3] M. Boulvin, Alain Vande Wouwer, R. Lepore, C. Renotte, Marcel Remy , *Modeling and Control of Cement Grinding Processes,* Control Systems Technolgy, 2003, ISSN 1063-6536.
[4] F. Cus, U. Zuperl, J. Balic, *Combined feedforward and feedback control of end milling system*, Journal of Achievements in Materials and Manufacturing Engineering, Vol. 45, Issue 1, pp. 79-88, 2011.
[5] Shue Li, Feng Lv, *Feedforward Compensation Based the Study of PID Controller*, Advances in Intelligent and Soft Computing, Volume 149/2012, pp. 59-64, 2012.
[6] L. Bruzzone, Rezia M. Molfino, *Experimental assessment of PID control with feedforward compensation of linear motors*, Proceedings of the 26th IASTED International Conference on Modelling, Identification, and Control, ACTA Press Anaheim, CA, USA, 2007.
[7] V. Santibanez, Rafael Kelly, *PD control with feedforward compensation for robot manipulators: analysis and experimentation*, Robotica (2001), pp. 11-19, Cambridge University Press, 2001.
[8] C.R. Costea, M. Abrudean, H.M. Silaghi, M.A. Silaghi, *Control of Flow Rate with Fuzzy Logic for Ball Mill,* Proceedings of 2010 IEEE International Conference on Automation, Quality and Testing, Robotics, THETA 17th edition, May 28-30, 2010, Cluj-Napoca, ISBN 978-1-4244-6722-8, pp. 153-156, 2010.
[9] G.C. Goodwin, S. Graebe, Mario E. Salgado, *Control System Design. Industrial Applications of Feedforward Control*, pp. 227, Valparaiso, Chile, 2000.
[10] H. Qian, X. Wang, H. Yu, *Classification and Recognition of Detecting Parameters for Cement Mill,* 2008 Second International Symposium on Intelligent Information Technology Application, ISBN: 978-0-7695-3497-8.
[11] http://en.wikipedia.org/wiki/PID_controller
[12] http://www.datatranlabs.com/category_s/37.htm
[13] M. Fuerstenau, K.Han, *Principles of mineral processing*, Society for Mining, Metallurgy, and Exploration, Inc., Littleton, 2003.
[14] A. Gill, *Calculating clinker grindability using an automated laboratory and computer technology*, Portland Cement Association Fall Technical Session, Oak Brook, IL, 1999.
.[15] A. Juhasz, L. Opoczky, *Mechanical activation of minerals by grinding: pulverizing and morphology of particles*, Ellis Horwood Limited Publishers, Chichester, United Kingdom, 1990.
[16] http://en.wikipedia.org/wiki/MATLAB
[17] http://en.wikipedia.org/wiki/Simulink
[18] S. Kawatra, *Advances in comminution*, Society for Mining, Metallurgy, and Exploration, Inc., Littleton, CO, 2006.
[19] J. Gutirrez, *Proportional-integral-derivative explained*, EE Times India, 2008.
[20] A.Blasco, *Particle size analysis reduces cement manufacturing costs,* Process Engineer,2008.
[21] J. Su, X. Zhang, X. Zeng, *Design and Simulation of Robust Ball Grinding Mill Control System*, 21-st Chinese Control and Decision Conference, Vol. 1-6, Proceedings, 2009.

Novel current monitoring techniques without shunt resistors

DACHIN Tudor[1], MEZA Serban[2], NEMES Marian [3], VODA Adriana[4], BADILA Florin[5]
[1] Department of Computers and Electrical Engineering,
"Lucian Blaga" University of Sibiu, Faculty of Engineering,
Tudor.Dachin@ulbsibiu.ro

[2]Technical University of Cluj-Napoca,
Faculty of Electronics, Telecommunication and Information Technology,
Serban.Meza@com.utcluj.ro

[3]Continental Automotive Systems S.R.L, Sibiu
Str. Salzburg nr.8, 550018, Marian.Nemes@continental-corporation.com

[4]Artsoft Cluj Napoca,
Adriana.Voda@artsoft-consult.ro

[5]Wenglor Electronic S.R.L. Sibiu
Str. Caprioarelor nr.2 550089, Florin.Badila@wenglor.com

Abstract **– *Current measurement for automotive electrical actuator applications (with motors or valves) is necessary for appropriate control in many cases and a safety requirement in all cases: the control algorithm may be dependent on the data but safety relevant functions will use it to determine possible over-current, over-temperature or failure conditions. This paper proposes an alternative method of monitoring the current, without using sensors or current shunts. Instead, measurements are made on the motor in the development stages and low/high frequency variations in the supply line are monitored, through low/high-pass filters, by available AD channels in the system. This results in cost reduction for the final product, by reducing hardware complexity.***

Keywords: ***measurement, shunt-less, EMC***

I. INTRODUCTION

Modern automotive systems often use actuators to change real-world process parameters, manipulate other objects or control another type of energy source (e.g. hydraulic pressure). When using electro-mechanical actuators (e.g. electro-valves) in an automotive application it is essential to monitor the voltage and/or current applied to them in order to properly control the desired output and keep the actuator in proper functioning condition. Most applications rely on having a (quasi-)constant voltage supply and monitor the current through the actuator.

Current measurement for automotive electrical actuator applications is necessary for appropriate control in many applications and is an important safety requirement in all cases: the control algorithm may be dependent on the data, but safety functions will use it to determine over-current, over-temperature or possible future or existing failure conditions. In order to monitor a current flow many ingenious methods have been developed, measuring it either directly (e.g. measuring the voltage over a current shunt) or indirectly (e.g. measuring the magnetic field intensity near the conductor). All precise measuring methods known today require the use of one (or more) elements (discrete circuitry or integrated devices) together with the actuator. This adds to the cost of the project in terms of hardware, software and development time.

This paper proposes an alternative method of monitoring the current, without using sensors or current shunts. Instead, measurements are made on the motor in the development stages to describe the start-up, normal operation and failure cases. Depending on the results, additional to regular monitoring, specific high frequency variations on the supply line are monitored, through high-pass filters, by available AD channels in the system and processed together with other monitored information. This is similar to sensory data fusion [1], but there is no discrete element/component that can be called "the sensor". The results are tangible cost improvements for the final product.

II. STATE OF THE ART

While current can be measured using Hall sensors [2] or current transformers, non-invasively, or with special transistors that have a sense terminal [3], current measuring techniques employed in automotive applications, like in other fields, use low resistive current shunts as they present a lower cost (with advantages and disadvantages [4,5]). Current shunt resistors are specifically made for current measuring applications and require small valued resistance, good

thermal stability and precise tolerance, in industry standard packages. Two problems with which engineers are faced are the power dissipation on the shunt (must be limited, not to destroy the part) vs. useful signal and the physical size of the shunt vs. the available layout space and product volume.

The needed measuring circuit is not composed only of the shunt itself, but also of a signal processing network: an analog filtering stage (which must be designed to "clean" the input signal from unwanted high frequencies while allowing as much useful signal to pass through), a voltage amplifier (high speed, precision operational amplifier or instrumentation amplifier) and an analog or digital controlled loop that uses the magnitude of the current to control the evolution of the system or shut it down in case of failure. The monitoring and control are usually integrated into a smart driver circuit. These additional components increase the cost of the product and the development effort.

Due to the competitive environment of automotive part manufacturers, there is a constant need for developing high quality but low cost products. In order to have a lower cost product, considerable research is done into discovering and adapting methods of "knowing" the current without having a shunt resistance or the smart driver with current measuring capabilities.

This research started by looking back at the basic operation of motors and how to drive them [6,7,8,9,10]. All motors generate the well known "**B**ack **E**lectro-**M**otive **F**orce" (BEMF) that can be used to monitor the motor speed and position.

If the precise position/rotation of the motor is needed, specific sensor shall be used based on speed, control type and budget. If there are multiple sensors in the system that may offer a feed-back path from the final controlled parameter, then the motor sensors may be eliminated and motor conditions can be inferred from the evolution of other parameters.

By reading BEMF at specific time points (start of actuation or change in actuation) many of the components mentioned previously can be deleted from the bill of materials. But the more information needed from the motor, the more complex the software implementation becomes and simply inferring values leads to false control. If the algorithm becomes too complex and unstable then real sensors are required and cannot be avoided.

III. OPERATING PRINCIPLE

The preliminary analysis discussed here and the proposal to use supply voltage variations (normally filtered out of the system) are the results of an observational study performed on specific linear actuators, that can be extended to any type of motor. The analysis started with a custom hardware implementation of the simple motor control circuit architecture below (Figure 1):

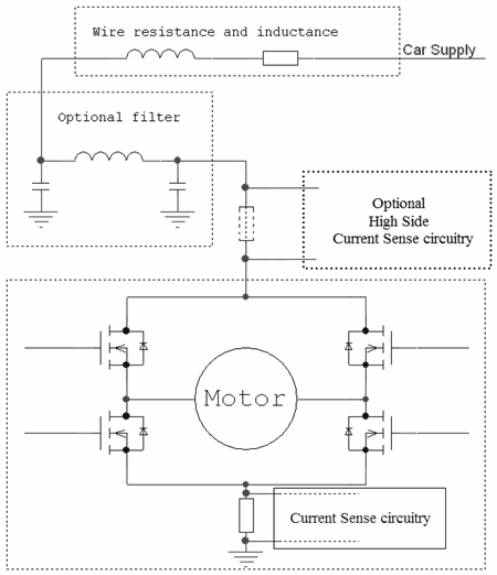

Figure 1. Reversible motor control with filtering and sense circuitry.

The motor is controlled through an H bridge, with low or high side current sensing, with simple ON/OFF, PWM or hybrid control. Based on the magnitude of the current and EMC regulations, the supply of the motor may have filtering circuitry for blocking high frequency harmonics. The entire circuit is connected with a real cable to the battery and chassis, which exhibits parasitic resistance and inductance. The current sense shunt will develop a voltage drop as current increases through it. This signal must be filtered, as it is considered "noisy" and too fast for regular control algorithms. After filtering, it is fed to the current sense circuitry, which adapts the level to the one needed by the next control loop block. Because the battery output resistance, the cable resistance and the motor/motor control circuit equivalent resistance form an electric circuit, the instantaneous "constant" current will generate voltage drops on them. The proposal is to eliminate the shunt resistor and adjacent circuitry (Figure 2) and harness the existing voltage variations:

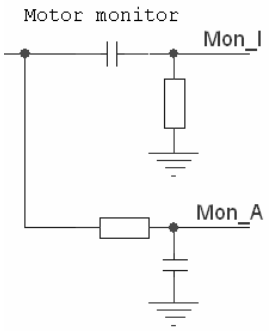

Figure 2. Indirect current measurement circuit.

Changes in current can be seen as changes in supply voltage, especially when measuring closer to the motor.

The problem with these changes is that they manifest as spikes in current and drops in voltage. They are considered unwanted and with proper filtering and local supply decoupling we try to eliminate them, which cannot be perfectly achieved.

IV. MEASUREMENTS

The current shunt circuit should be replaced with a simple yet effective low pass / high pass filter pair, connected to free A/D (analog to digital) input channels of the system micro-controller. The sampling speed must be sufficient to cover the variations of the supply, caused by the motor operation. By interpreting the ripple on the supply voltage, with an appropriate software interpretation and secondary data inputs from existing sensors connected to other systems (be it in the same electronic control unit or in other controller linked through communication networks) the current can be estimated with sufficient accuracy for safety and/or control function.

The measurements have been performed using a HIL device in order to have the same testing conditions or the different tests. The data acquisition was made by using an oscilloscope. The tested system consists of a linear motor with various loads applied, to mimic real-life operating conditions:

- unloaded (free-running motor; possible failure);
- physical limitation at half excursion (one of the intended applications of the motor);
- physical limitation at beginning of movement (stuck motor; possible failure);
- nominal load over the entire excursion range.

In the first case, we control the bridge with a PWM signal (0-13V) and measure the average voltage needed for driving the motor [11] with a specific current (average value drawn from the power supply). For the specific motor the datasheets prohibits currents higher than 12A, so that was the superior limit for testing. To plot the motor voltage-current dependency, 1A steps from 1A to 12A were used. The first test was performed with unloaded motor; afterwards the rest of the above mentioned loading scenarios were tested. The resulting data was plotted in Figure 3 below (showing the MON_A parameter from Figure 2):

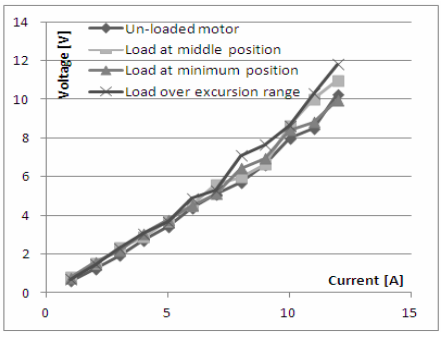

Figure 3. Average current drop on motor and control circuitry.

The average current (monitored like in Figure 2) through the motor and control circuitry is dependent on many parameters (temperature drift, part tolerance, ageing, and mechanical manufacturing tolerances etc.) and will vary from product to product. Statistical analysis must be performed on a large sample batch before deciding if the monitoring accuracy is enough for control or simple over-current detection. Statistical data gathering and statistical simulations are necessary and calibration routines during product final assembly and/or during normal operation can reduce offset errors. A temperature sensor (low cost and common in many applications) can be used to normalize the measured values, by knowing the temperature dependency of the motor's resistance.

On the other hand, disturbances in the supply line were measured to see if there is any correlation between initial transient voltage amplitude and drawn current. The same defined loading conditions were applied and the amplitude of the voltage transients (measured as maximum positive peak to stable voltage) versus desired current draw plotted. For this, the power supply was current limited to a specific value and then turned on. The result is plotted in Figure 4 below:

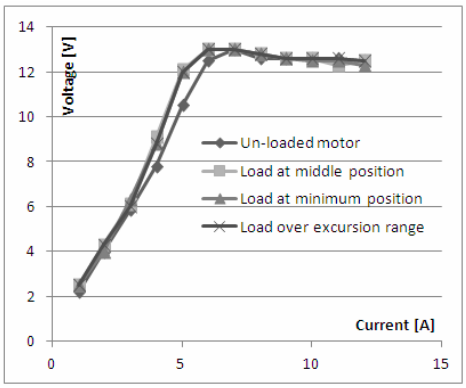

Figure 4. Peak-to-stable transient voltage amplitude at start-up.

The voltage dependency is almost linear up to a certain current value (5-6A), after which it seems to saturate (voltage excursion is limited by maximum supply voltage), showing that this specific type of motor could be used with constant voltage control and inferring the current value from the initial transient and adaptive look-up tables, the average voltage on the motor and the temperature of the motor. Other motors will surely behave differently, but the alternative of not using shunts or other measurement types and relying on motor parameter calculations is valid. This shifts the production costs over to software development, but reduces long-term costs as fewer and lower price parts are necessary.

Voltage control (using DC-DC converters and efficient filtering) is gaining ground in currently under development applications because of more stringent NVH (**N**oise **V**ibration **H**arshness) requirements from

the customer. Also, less EMC related qualification issues are likely to occur without high-frequency switching. This is the case for controlling analog valves (and avoids clicking/hammering sounds) or driving motor with long connecting wires (coupled with high current switching they will become emitting antennas). Because the control is linear and switching does not have to occur continuously, a learning algorithm can be implemented to measure the desired effect and correlate it with temperature, transient voltage, average measured voltage and desired current trough the actuator. Efficiency is maintained in a wide range of operation voltages by using DC-DC converters, but only if the actuator has a low start-up voltage. For these applications the following measurements might be relevant: settling time for a step change in supply voltage, from no supply to a given voltage value (derived from current requirements by the software component of the control unit). For the tests related to Figure 4 the settling time was also measured and the data plotted in Figure 5 below:

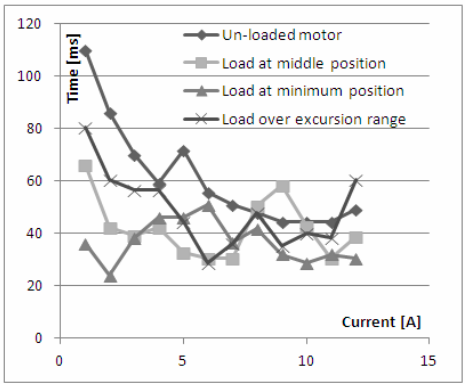

Figure 5. Transient voltage duration before steady state is reached.

At first sight there seems to be little information to take from the plot, as the current-settling time curves are non-linear and non-monotonous; but some additional information is available at a glance: settling time in case of unloaded motor is clearly higher, providing another way of detecting unconnected load failures. Initial loading of the motor in case of the "load at minimum position" test case indicates shorter transient times. The main problem is that there are no clear ranges to be defined in order to determine a specific condition. Therefore linearization of the transient duration measurement is needed in order to make information gathering valid, at least on a range of the operating conditions if not possible over the entire datasheet motor operating range. This has been tried in Figure 6.

At least for the specific motor tested, linearity over a wide enough interval is not possible by plotting transient duration vs. transient amplitude.

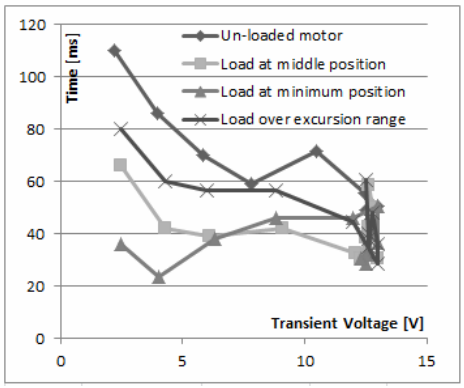

Figure 6. Transient voltage duration plotted as a function of transient amplitude, demonstrating high non-linearity.

A different parameter or set of parameters should be used in order to extract additional information.

V. CONCLUSIONS

The measurements prove that it is possible to acquire current information from the voltage on the load and from the amplitude and/or duration of transients. The hardware implementation is therefore simplified by using low cost filtering circuitry and avoiding high cost application specific parts. The software complexity and resource requirements increase with control accuracy requirements. For some applications, it is possible and feasible to use this simplified hardware interface, if cost reduction or hardware minimization is needed.

REFERENCES

[1] H. B. Mitchell, Multi-sensor Data Fusion – An Introduction, Springer-Verlag, Berlin, 2007, ISBN 9783540714637.

[2] John Cummings, Michael C. Doogue, Andreas P. Friedrich, Recent Trends in Hall Effect Current Sensing, AN295045

[3] Current sensing power MOSFETS, NXP Semiconductors AN10322, 2009

[4] William Koon, Current sensing for energy metering, Comparison of various current sensing technologies, Analog Devices

[5] Bill Drafts, Methods of Current Measurement, Pacific Scientific - OECO, 2004

[6] Kristin Lewotsky, Linear Motors Part I: The Straight Scoop, 2008

[7] Alrifai, M., Zribi, M., Krishnan, R., et al., Nonlinear Speed control of Switched Reluctance Motor Drives Taking into Account Mutual Inductance. J. of Contr. Sc. Eng., ID 491625, Hindawi (2008)

[8] Astrom, K.J., Wittenmark, B., Adaptive Control, Addison-Wesley, New York (1989)

Location Aware Control — A Merchant Shipping Perspective

DAVE Dhiren, NALBALWAR Sanjay, GHATOL Ashok
Dr. Babasaheb Ambedkar Technological University, Lonere, India,
Department of Electronics & Telecommunication Engineering,
E-Mail: dpdave3564@gmail.com, nalbalwar_sanjayan@yahoo.com, vc_2005@rediffmail.com

Abstract *– Location aware management and location based automation are fast upcoming technologies which are facilitated by availability of reliable, accurate, and cheap global positioning systems and location based services. This paper presents a software approach, for location aware control, specifically with merchant shipping perspective. This leads to improved safety, increased automation, prevention of pollution and reduction in the work burden of the crew. The GPS is used in conjunction with the regionally accessible nested global shorelines (RANGS) dataset for demonstration of a case of proposed location aware software architecture. Two C++ classes developed towards implementation of this software have been discussed. As an example of shipping application, a control system mandatory on all oil-tankers above 150 GRT (and few other vessels) called ODMCS has been made location aware using the novel software developed and the results obtained have been critically examined.*

Keywords: *GPS, Location Aware Control, ODMCS, RANGS.*

Symbols and Abbreviations:

DOP	– Dilution of Precision (of GPS data)
GPS	– Global Positioning System
GRT	– Gross Tonnage
GSHHS	– Global Self-consistent Hierarchical High-resolution Shorelines
LBS	– Location Based Services
MARPOL	– Marine Pollution
MFC	– Microsoft Foundation Classes
NIMA	– National Imagery and Mapping Agency
NMEA	– National Marine Electronics Association
ODMCS	– Oil Discharge Monitoring and Control System
RANGS	– Regionally Accessible Nested Global Shorelines
UTC	– Coordinated Universal Time
WDB	– World Data Bank
WVS	– World Vector Shorelines (dataset)

I. INTRODUCTION

In last decade, the availability of low cost, reliable and accurate GPS [1] and concurrent growth of mobile telephone based LBS has paved the way for serge in location aware management and control systems. Such systems are finding acceptance in mobile applications like cellular telephony, rail/road based transportation, aerospace, navy, merchant shipping etc [2],[3]. This paper discusses the development of reusable C++ class templates, developed around MFC, which may be used for development of software for location aware management and automatic control, specifically in merchant shipping applications. It can be shown that such systems provide improved safety and reliability, increased automation, prevention of pollution and reducing the work burden of the crew[4]-[7].

A location aware system should be able to receive location inputs, maintain location database, and take location based actions after processing the inputs and location database.

The information about the current location may be received using various technologies like GPS, LBS, Blue-tooth etc. The location database may include important land-marks, geographical information, map etc. For marine applications, this may include shorelines, locations of ports, important routes, special areas, environmentally sensitive areas etc.

This information may be processed and used for various purposes like navigation, route planning, collision avoidance, ship-board related waste management, environment protection, anti-piracy measures etc.

Section II briefly describes NMEA protocol, a popular data communication standard used for communication of data between GPS and other ship-board devices and the state-machine implementation of the same. Section III gives a brief description of RANGS database. Section IV presents software implementation issues like features requirements, software architecture, simple flowcharts and templates for the two important classes namely CShorelines and CGPSInput, which act as the main data retrieval and processing mechanism. Section V describes implementation details. Section VI discuss the test results of a location aware oil discharge monitoring and control system developed based on these classes.

II. NMEA 0183 STANDARD

The GPS is a worldwide radio-navigation system formed from a constellation of 24 satellites and their

ground stations. Most GPS devices onboard the ships support NMEA protocol for data communication between GPS device and the computer.

The present work is based on the NMEA 0183 specification [8]. The NMEA 0813 standard for interfacing marine electronics devices specifies the NMEA data sentence structure as well as general definitions of approved sentences. However, the specification does not cover implementation and design. The author has presented the software design tasks needed to parse through NMEA sentences robustly. The technique for parsing and data integrity checking is proposed.

NMEA data is sent in 8-bit ASCII where the MSB is set to zero. The specification also has a set of reserved characters. These characters assist in the formatting of the NMEA data string. The specification also states valid characters and gives a table of these characters ranging from 0x20 to Ox7E. Prefix 0x indicates that number is hexadecimal.

As stated in the NMEA 0183 specification version 3.01, the maximum number of characters shall be 82, consisting of a maximum of 79 characters between start of message "$" or "!" and terminating delimiter <CR><LF> (0x0D and 0x0A). The minimum number of fields is one. The basic format for NMEA sentence is as below:

$aaccc,c--c*hh<CR><LF>

Where the various characters/fields represent the following,

$ - Start of sentence
aaccc - Address field/Command
"," - Field delimiter (0x2C)
c--c - Data sentence block
* - Checksum delimiter (0x2A)
hh - Checksum field (the hexadecimal value represented in ASCII)
<CR><LF> - End of sentence (0x0D 0x0A)

III. RANGS DATASET

To make a software to be used in marine applications location aware, at the very minimum, it requires the information on present location (for e.g. from GPS) as well as information on shorelines in vector format and on a global basis.

The WVS provides data on shorelines in vector format on a global basis. The WVS was originally provided by the National Imagery and Mapping Agency, USA. In fact, an improved version of WVS called RANGS [9] has been used for the purpose of present work. RANGS files are based on the GSHHS files. The GSHHS dataset had been derived by Wessel and Smith from the WVS dataset combined with additional WDB dataset [10]. Wessel and Smith also developed various low resolution versions from the main WVS dataset using the algorithms developed by Douglas and Peucker [11].

The RANGS dataset is organized into cells of 1° longitude x 1° latitude covering entire globe. This forms 64,800 cells. The cell contains an array of points (lon,lat) representing the shoreline segments, forming a simple, closed cell polygon. It also includes the inland lakes and ponds. The dataset also contains information on whether the inside of a particular polygon is land or water.

RANGS files are available in five resolution versions (full resolution = 0.1, high = 0.2, intermediate = 1.0 , low = 5.0 and crude = 25 km). The corresponding files are named as rangs(0), rangs(1), rangs(2), rangs(3) and rangs(4) respectively. The dataset for each resolution level n ($0 \le n \le 4$), is organised in group of three files.

(i) Cell Address Table file, Rangs(n).cat
(ii) Cell Extraction List file, Rangs(n).cel
(iii) Shoreline data file, gshhs(n).rim

The details of the structure of these files and how they are organised to access RANGS information may be found in [9]. The flowcharts for reading and interpretation of RANGS data is given in section V.

IV. SOFTWARE DESIGN

Typical location aware software for application in merchant shipping is expected to cater for any one or more of the location based features like alarm, monitoring and control, management etc. Accordingly, the specific features to be taken into consideration while developing such a software are as follows-

(i) Ability to receive location information from various location providers (GPS, Blue tooth devices, Mobile Networks etc.), in real-time, using commonly used communication protocols like RS232C, USB etc.
(ii) Ability to interpret/process location data format for extraction of information like position, speed, heading etc.
(iii) Ability to import shoreline data in vector format.
(iv) Ability to integrate into other applications easily.
(v) To calculate distance between any two points on surface of earth along the great circle.

To calculate the distance between any two points, represented by say, P1($\theta 1, \delta 1$) and P2 ($\theta 2, \delta 2$), where $\theta 1$ and $\theta 2$ represent longitudes and $\delta 1$ and $\delta 2$ represent latitudes, various methods for e.g. Haversine's Formula and Spherical Law of Cosines [12],[13] etc are available.

This paper proposes using spherical law of cosines because of its simplicity and accuracy, which is good enough for many applications in shipping. The spherical low of cosines is as follows.

$$D = \text{acos}(\sin(\delta 1) \times \sin(\delta 2) + \cos(\delta 1) \times \cos(\delta 2) \times \cos(\theta 2 - \theta 1)) \times R \quad (1)$$

Where, D is the distance (in Knots) and R is the mean radius of earth (km). This formula gives the results to the accuracy of about one meter [14].

V. IMPLEMENTATION

Two of the most important classes which encapsulate most of the data and functionality of a location aware system, for the proposed shipping applications, are

implemented in native C++ using Microsoft® Visual C++ 6.0 integrated development platform.

CShorelines class:

The CShorelines class is designed to encapsulate data and functions to meet the following main application requirements.

(i) Ability to import (read, interpret and display) shoreline data in vector format.
(ii) Save/Load the display image to/from storage media (hard-drive).
(iii) Dynamically display longitude and latitude at any user specified point anywhere on map.
(iv) To measure distance between any two points on the map in nautical miles.
(v) Zoom in/Zoom out whole map.
(vi) Zoom-in a selected portion of the map for closer inspection.
(vii) Determination of point background (ground or ocean) programmatically.
(viii) Determine if a given position specified by (longitude, latitude) lies within a special area [5],[7].
(ix) Calculate the distance of a point from the nearest shoreline.
(x) Configurable with different display resolutions and pallets.

CGPSInput class:

The CGPSInput class is designed to encapsulate data and function to meet the following application requirements.

(i) Ability to receive location data (from GPS) using COM/USB port or blue tooth device into the application.
(ii) Ability to read, interpret and process NIMA sentences received from GPS device.
(iii) Ability to handle timeouts and other data transmission errors (related to GPS).
(iv) Display the important information received from GPS device on the monitor of computer.

The header and implementation files for CShorelines and CGPSInput classes may be obtained from the author on request.

V.1. NMEA interpreter state machine implementation

Fig. 1 shows the state machine implementation of a typical NMEA sentence interpreter. The state machine tracks the protocol state and any errors that may occur during the data transfer. This approach allows to track the state of the system (within the protocol) and also to recover from errors like timeouts, checksum errors etc.

The various states of the state machine are described below:

(i) State SOS (Start of Sentence): In this state the state machine looks for the '$' (0x24) that is start of the sentence character.
(ii) Receive address/command: In this state, the state machine collects characters until it receives a ',' (0x2C) character. The variable length address field allows parsing of any undefined or proprietary sentences.
(iii) Get sentence data: In this state, the state machine continues to collect data and also performs a checksum until it receives a checksum delimiter "*" (0x2A) or sentence terminator <CR><LF> (0x0D 0x0A).
(iv) Get checksum character (First): In this state, the state machine simply waits for the arrival of first checksum character.
(v) Get checksum character (Second): In this state, the machine waits for the arrival of second checksum character. After receiving the second checksum character, the received checksum is verified against the calculated checksum.
(vi) Get sentence terminator (ST), first character: In this state, the state machine simply waits for the first sentence terminator character (0x0D).
(vii) Get sentence terminator (ST), second character: In this state the state machine waits for the arrival of 0x0A character. When this character is received, sentence is complete and may be processed by NMEA interpreter.

The timeout transition in the majority of the states is necessary so as to synchronize the state machine when no data is received for a period of time. The state machine makes a transition to initial state, when no data has been received for a period of time (timeout). The timeout duration is application dependent. A 4800 baud device typically sends data every 1 to 2 seconds and timeout of 3 to 4 seconds is sufficient.

The NMEA processor receives NMEA sentences from the state machine and processes this sentence for extracting information. The state machine receives the data through the RS-232C serial port of the computer.

The author has used the following commands in the application developed [8].

GPGGA - Global Positioning System Fix Data
GPGSA - GPS DOP and Active Satellites
GPGSV - GPS Satellites in View
GPRMB- Recommended Minimum Navigation Information
GPRMC- Recommended Minimum Specific GPS/TRANSIT Data
GPZDA - UTC Date / Time and Local Time Zone Offset

V.2. Flow chart for reading and interpretation of RANGS data

The software, for reading, interpretation, and rendering of RANGS data, is developed using native C++ code. Figures 2(a) to 2(c) depict the flowchart for reading and interpretation of the RANGS data.

V.3. Algorithm to detect if the ship is in special area

Following procedure is used to detect the presence of

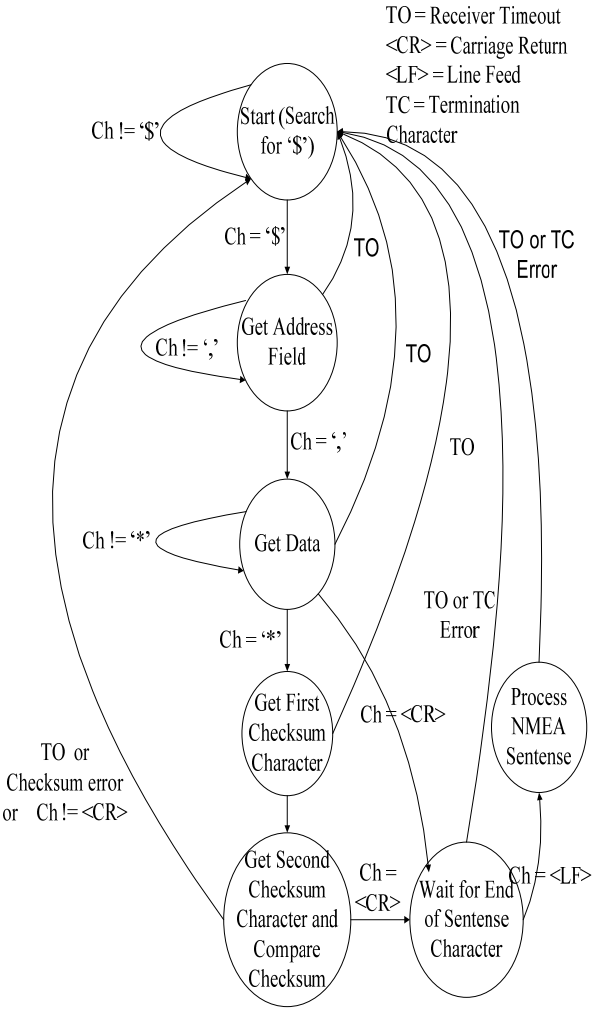

Figure 1. The state machine implementation for receiving GPS data.

ship in any of the special areas as defined in MARPOL regulations[16].

1. Get ship position (Longitude, Latitude).
2. Start with any one of the special areas.
3. Get array of points forming a enclosing polygon (for selected special area). Each point in array represents a pair of longitude and latitude.
4. Is ship within selected special area (polygon)? If yes, go to step 7.
5. All special areas checked? If yes, go to step 8.
6. Select next special area. Go to step 3.
7. Set the ship is in restricted area alarm. Return from Procedure.
8. Clear the flag representing alarm that the ship is in restricted area. Return.

V.4. Algorithm for finding the nearest land

Please see [16] for the definition of 'nearest land'. The algorithm is as follows-

1. Current_Ship_Position = (Longitude_From_GPS, Latitude_From_GPS).
2. Cell_Longitude = Longitude of the lower left corner of the current cell.
3. Cell_Latitude = Latitude of lower left corner of current cell.
4. Longitude = (Cell_Longitude - 1). If Longitude is -181, Longitude = 179.
5. Latitude = (Cell_Latitude – 1). If Latitude < 89 OR Latitude > -90 Than Cell_Latitude = Latitude-1.
6. Get the cell, whose lower left corner has longitude and latitude one less than that of the current cell.
7. Get the array of points for the cell obtained in step 6 and check if any of the points in the array has a distance of 50 Nautical miles or less, from the current position of ship, using equation (1). If yes, set the flag for this alarm and terminate.
8. Otherwise, repeat step 4 to 7 for all the cells up to one longitude and one latitude greater than the current cell (total 9 cells including current cell).
9. If all the cells are checked, clear the flag for corresponding alarm and terminate.

VI. RESULTS AND CONCLUSIONS

The testing of CGPSInput class was carried out using GARMIN® GPS-72™ device, giving accuracy of about 15 meters, 95 % times [17], which was interfaced with an IBM-PC compatible computer with Intel® Core2 Duo™ processor running Windows® XP™, using RS232C interface cable. The CShorelines class was developed and tested around Rangs(0).cat, Rangs(0).cel, and Rangs(0).rim files, giving a resolution of about 100 meters [9].

A control system mandatory on all oil-tankers above 150 GRT (and few other vessels) called ODMCS [5] has been made location aware using the novel software developed.

The developed system automatically detects the presence of ship in special areas[4],[5],[16], generates alarm and takes appropriate control actions. Also, it generates alarms and initiates suitable control actions when the ship is less than 50 nautical miles from the nearest shore, in accordance with the MARPOL regulations.

It is not possible to publish the detailed results of the developed system (ODMCS) because of limit of space. However, the complete data generated may be obtained from the author on request.

A patent application has been filed by the author on the same subject [2]. The complete specifications for the patent are also published by the Indian patent office. The same is available on the website of the Indian patent organisation.

More details on the hardware development for this work may be found in [5],[7]. The simulation of the ODMCS was done as a separate work by the same author and the details on the same may be found in [4].

ACKNOWLEDGEMENTS

The help and support from Tolani Maritime Institute, Induri (India) is gratefully acknowledged.

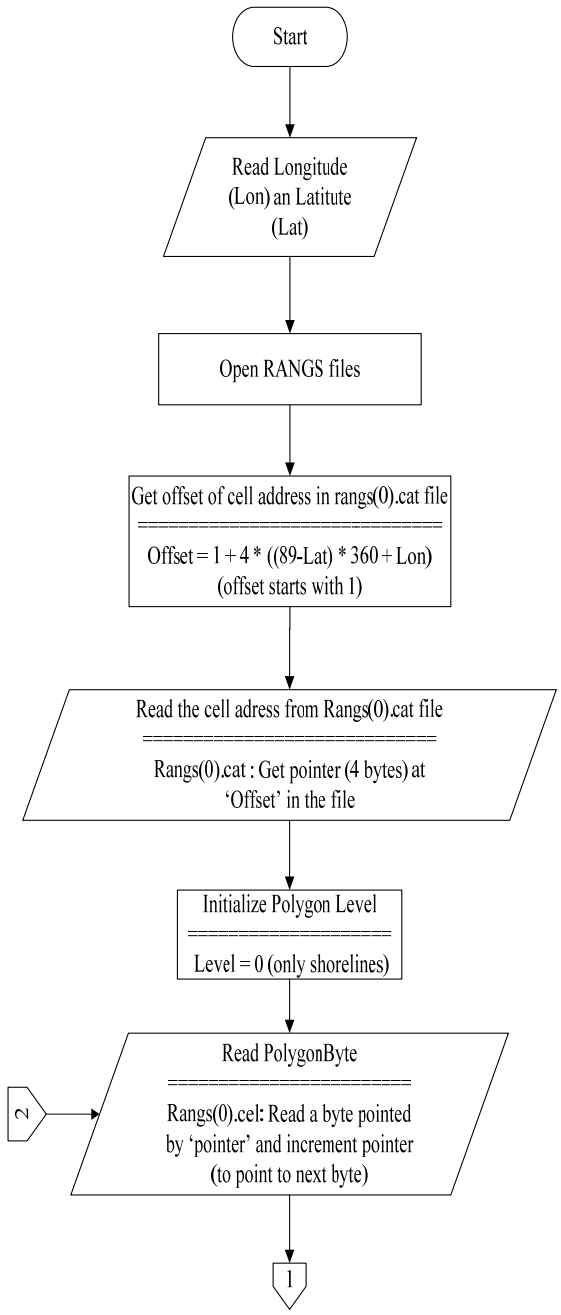

Figure 2(a). Flowchart for reading RANGS data.

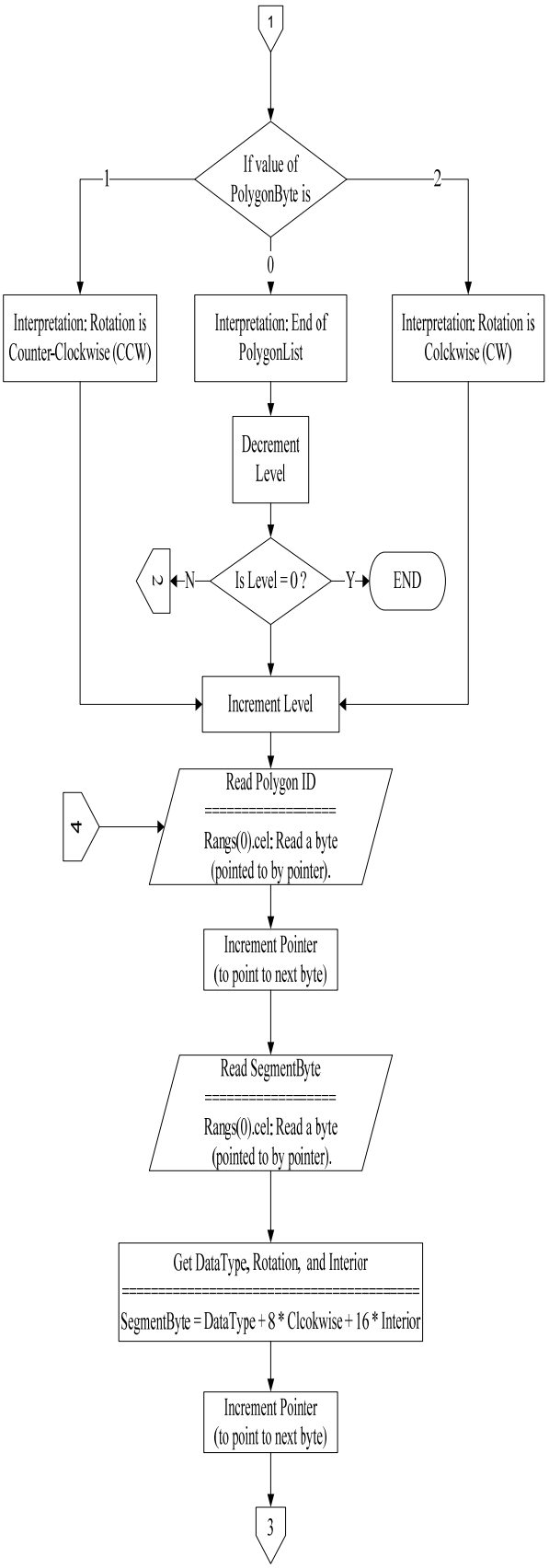

Figure 2(b). Flowchart for reading RANGS data.

REFERENCES

[1] __, Global Positioning System. Available Online: http://en.wikipedia.org/wiki/Global_Positioning_System, accessed on 17th Nov, 2011.

[2] D. Dave, A. Ghatol, Patent Application: "*Location Aware Pollution Control For Ships*," Application Number: 2019/MUM/2009, Dated 03/09/2009.

[3] J. Nord, K. Synnes, and P. Parnes, "*An Architecture for Location Aware Applications*," Proc. 35th Hawai Int. Nat. Conf. on System Sciences. 2002. Available Online: http://csdl2.computer.org/com/proceedings/hicss/2002/1435/09/14350293.pdf

[4] D. Dave, and A. Ghatol, "*A Software Paradigm for Complete Automation of ODMCS,*" Int. Nat. Conf. on All Electric Ship, London, 2007.

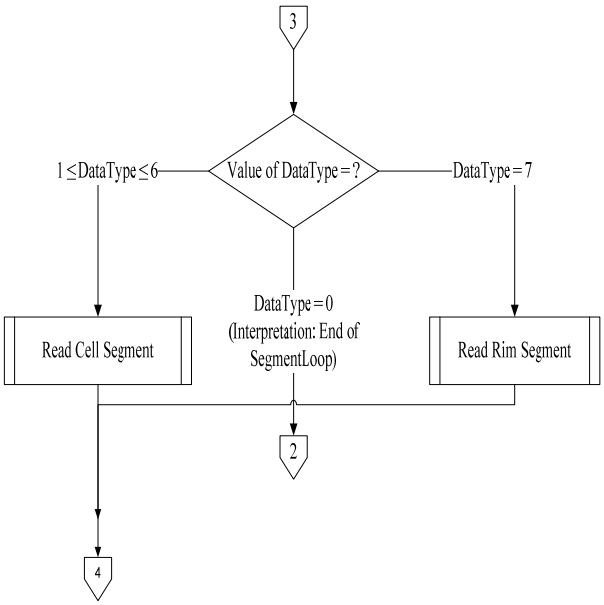

Figure 2(c). Flowchart for reading RANGS data.

[5] D. Dave, "*Design of Completely Automatic ODMCS,*" International Journal of Maritime Engineering, Vol. 152(A3), pp. 147-157, 2010.

[6] D. Dave, "*Location Based Intelligent Pollution Control for Ships,"* Int. Nat. Maritime Tech. Conf., Sept 20, Mumbai, India, 2010.

[7] D. Dave, A. Ghatol, "*Software Architecture for Modeling, Simulation and Automation of Ballast Water Discharge from Oil Tankers,*" International Journal of Engineering Simulation, Vol. 10(2), pp. 27-35, 2009.

[8] National Marine Electronics Association, "*NMEA 0183 Standard For Interfacing Marine Electronic Devices*," Version 3.01, January 1, 2002.

[9] R. Feistel, "*New Shoreline Map-Drawing Data Available,*" Eos Transactions, American Geophysical Union, Electronic Supplement, Vol. 80(22), pp. 249, 1999. doi:10.1029/99EO00188.

[10] P. Wessel, and W. H. F. Smith, "*A Global Self-consistent, Hierarchical, High-resolution Shoreline Database,*" Journal of Geophysical Research, Vol. 101(B4), pp. 8741-8743, 1996.

[11] D. Douglas, and T. Peucker, "*Algorithms for the reduction of the number of points required to represent a digitized line or its caricature*," The Canadian Cartographer, Vol. 10(2), pp. 112–122, 1973. doi: 10.3138/FM57-6770-U75U-7727.

[12] R. Sinnott, "*Virtues of the Haversine*," Sky and Telescope, vol. 68(2), pp. 195, 1984.

[13] W. Gellert, S. Gottwald, M. Hellwich, H. Kästner, and H. Küstner, *The VNR Concise Encyclopaedia of Mathematics*, 2nd ed., ch. 12, Van Nostrand Reinhold: New York, 1989.

[14] __, Calculate distance, bearing and more between Latitude/Longitude points. Available Online: http://www.movable-type.co.uk/scripts/latlong.html accessed on 17th Sept, 2011.

[15] D. Kruglinski, S. Wingo, and G. Shepherd, Programming Microsoft Visual C++, Microsoft Press, 1998.

[16] International Maritime Organisation, MARPOL 73/78. Mumbai: Bhandarkar Publications, pp. 58-65, 2002.

[17] GARMIN GPS-72 Specifications. Available Online: http://www8.garmin.com/products/gps72/spec.html, accessed on 17th Nov. , 2011.

Depth Map Calculation for Autostereoscopic 3D Display

HROZEK František, IVANČÁK Peter

Technical University of Košice, Slovak Republic,
Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics,
Letná 9, 042 00 Košice, Slovak Republic, E-Mail: frantisek.hrozek@tuke.sk, peter.ivancak@tuke.sk

***Abstract* – Creation of content for 3D displays is very actual problematic. This paper focus on this problematic and is divided into two parts. First part presents various 3D displays and displaying technologies, especially stereoscopic displays – passive, active and autostereoscopic. Second part presents application that calculates depth map from stereoscopic image and was developed at DCI FEEI TU of Košice (Department of computers and informatics, Faculty of electrical engineering and informatics, Technical university of Košice. This application use three alghoritms for depth map calculation: Block Matching, Graph Cut and Semi-Global Block Matching. Depth map calculated in this application is used as part of the input image for autostereoscopic 3D display Philips WOWvx.**

Keywords: *depth map calculation, passive autostereoscopic display, 2D-plus-depth, Philips WOWvx.*

I. INTRODUCTION

3D display technologies have been widely used in architecture, engineering, education, entertainment, etc. Compared to 2D display, 3D provides additional depth information and represents objects in a more natural and accurate way. There are several technologies that can be used for 3D display [1]. The most common ones are stereoscopy, volumetric visualization and holography. Research at DCI FEEI TU of Košice is focused on stereoscopic technologies.

Stereoscopy is based on the way human brain perceives the surrounding objects. When a person looks at an object, it is seen by each eye from a slightly different angle. Human brain processes the information and enables stereoscopic vision. A stereoscopic image (or a video) is a composition of left and right images captured by the two cameras. The image is then presented to the viewer in such manner that each eye can see its image (left eye see left image and right eye see right image). In this way human brain interprets a 2D image as a 3D scene.

This paper presents application that calculates depth map from stereoscopic image or video using three algorithms (Bloch Matching, Semi-Global Block Matching and Graph Cut). Calculated depth map is used as part of the input image for autostereoscopic 3D display Philips WOWvx which allows users to see 3D images (video) without glasses or any headgear.

II. RELATED WORK

Currently exists several applications which calculate depth map. Some of them are:

- *Depth map creator* – depth map is calculated from stereo pair (left and right image). `http://www.3dphotopro.com/soft/depthmap/help.htm`
- *2D-to-3D convertion* – this software is available as a set of plug-ins for Adobe After Effects CS5 with partial support for Nuke 6.2. Software can use for depth map calculation three methods called: Depth from Motion, Depth from Focus and Depth Effects — depth from geometry. `http://www.yuvsoft.com/products/2d-to-3d-conversion/`
- *StereoTracer* – depth map is calculated from stereo pair or can be drawn manually by user. `http://triaxes.com/products/stereotracer/`
- *Ocula* – this software is available as a set of plug-in tools for Nuke. Input for this software is stereo pair or stereo video. `http://www.thefoundry.co.uk/products/ocula/`

III. 3D DISPLAY TECHNOLOGIES

3D displays use several technologies to create 3D image. Each technology has its advantages and disadvantages. 3D displays can be divided into these cathegories [2]:

- holographics displays
- volumetric displays
- stereoscopic displays
  - passive
  - active
  - autostereoscopic

Research at DCI FEEI TU of Košice is focused on stereoscopic 3D displays and their displaying technologies (passive and active stereoscopy, autostereoscopy).

A. *Active Stereoscopy*

Active stereoscopy is now probably the most common method used in 3D televisions. It has high picture quality, frame rate is doubled (120 Hz – 60 Hz for each eye) and also picture quality is very high. Disadvantage of these technology is that the user need for viewing special electronic glasses which are synchronized with the remote broadcast source and alternately shows images for the left and for the right eye.

B. *Passive Stereoscopy*

Passive stereoscopy use light polarization for separation of left and the right image. For viewing user needs glasses with polarization filters but these glasses do not include any active parts.

C. *Autostereoscopy*

Autostereoscopy is method on displaying stereoscopic images without the use of special headgear or glasses. There are two autostereoscopic methods: *active* (special camera track user's eyes position) and *passive* (no special hardware is needed).

Passive autostereoscopy use two most common technologies for displaying of 3D image: *parallax barrier* and *lenticular lens*.

Another one passive autostereoscopic method is 2D-plus-depth. This method was developed by Philips and 3D image is created by using 2D image and depth map. Input image example for autostereoscopic 3D display Philips WOWvx is shown in Fig. 1. This image contains from two sub images: 2D image and depth map. More details about this 3D TV at [3] and more details about 2D-plus-depth method at [4].

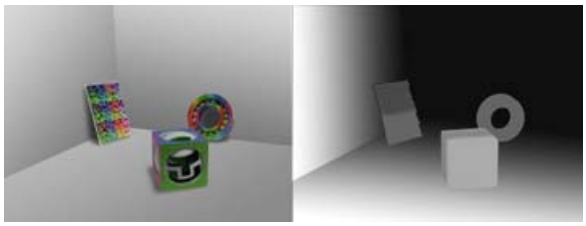

Figure 1. Input image example for Philips WOWvx: 2D image (left half) and depth map (right half).

IV. DEPTH MAP APPLICATION

This application was developed at DCI FEEI TU of Košice. Input for this application can be static image (left and right image / stereo image) or dynamic image (stereo video). Application use for depth map calculation these three algorithms:

1. *Block Matching (BM)* – is a way of location matching blocks in a sequence of digital video frames for the purposes of motion estimation.
2. *Semi-Global Block Matching (SGBM)* – is more accurate than standard BM and with almost the same calculation speed as standard BM. This algorithm tries to find the best possible alignment between two images.
3. *Graph Cut (GC)* – normal stereo matching algorithms try to match a pixel in the left image to a pixel in the right image based on some individual property like color. Although this is a fast and reasonably accurate process it does not deal with interlinear consistency. Graph Cut algorithm incorporates interlinear consistency and takes in account the interpixel properties like adjacency. This results in a much more accurate disparity depth map.

Interface of created application is shown in Fig. 2.

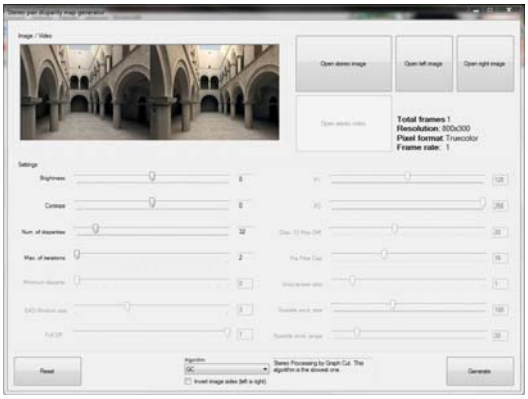

Figure 2. Interface of created application.

Creation of input image for Philips WOWvx in this software consists from these steps (Fig. 3):

1. load input file (left and right image, stereo image or video)
2. set parameters for calculation (e.g. brightness, contrast, number of iterations)
3. choose algorithm (BM, SGBM or GC)
4. generate image/video (left part of input image/video + calculated depth map)
5. use created image/video in Philips WOWvx

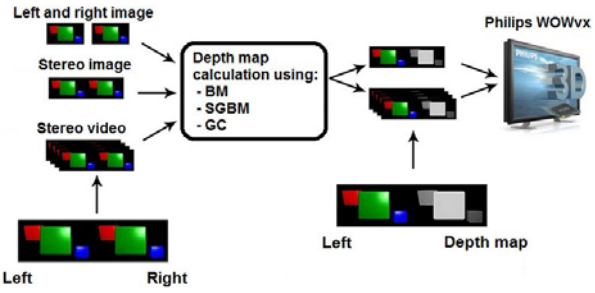

Figure 3. Creation process of depth map for Philips WOWvx.

V. GRAPH CUT ALGORITHM

This section in short describes GC, since it gives the best outputs (according to tests presented in section VI.) More details about GC algorithm and its optimization can be found for example in [5] [6] [7] [8].

Graph Cut Algorithm is based on graph representation $G = (V, E)$, where individual vertices $v \in V$ are pixels and edges $<v_i,v_j> \in E$ connects adjacent pixels. In the classic 2D image pixel can have 4 or 8 neighbor pixels. Each vertex v and edge $<v_i,v_j> \in E$ has a price on an object which will be segmented. Let the set of all image pixels is marked as I and set of all pixels $(p,q) \in I$ which represents neighboring pixels is marked as N. Let O be the set of all pixels which belong to the object and B is the set of all pixels which belong to the background. Let each pixel in the picture i_k also belongs into one of the binary classes $L_k \in \{O, B\}$, where O and B represent the set of object and background pixels. Then $L = (L_1, L_2, ..., L_{|I|})$ define the final segmentation.

Method can be initialized interactively or automatically by identification of one or more of the points which represents object or background.

To find the result is used function [4]:

$$C(L) = \lambda R(L) + B(L) \qquad (1)$$

To minimize the function $C(L)$ (1) is used a special type of graph $G_{ST} = (V \cup \{s, t\}, E)$. Hereto graph V (vertices of this graph correspond to the pixels of the image I) are attached two special end vertices s and t. For each graph vertex exist only one edge, which connects this vertex with end vertex s and only one edge, which connects this vertex with end vertex t. Edges E in the graph G_{ST} are classified into two categories: *n-edges* and *t-edges*. A set of *n-edges* connects a pair of neighbor pixels. Price of these n-edges is derived from the border member $B(L)$ (1). A set of *t-edges* connecting graph vertices which represents pixel of the image with end vertices. Price of t-edges is derived from $R(L)$ (Eq. 1). The s-t cut of the G_{ST} graph is set of edges, which this cut intersect and divide this graph into two disjoint subsets S and T, $s \in S$ $t \in T$ and there is no direct path that connect end vertices s and t.

Next example shows how GC cuts picture P (Fig. 4). Graph G created from picture P is shown in Fig. 5 and cut through this graph using GC is shown in Fig. 6.

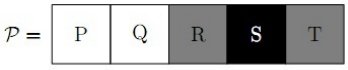

Figure 4. Picture P.

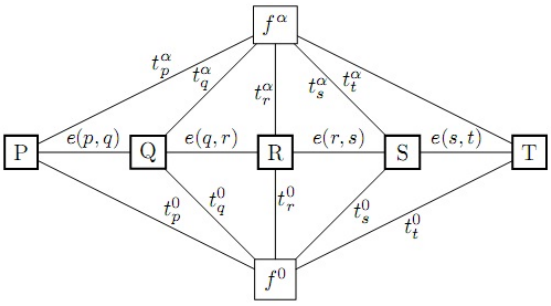

Figure 5. Graph G created for picture P.

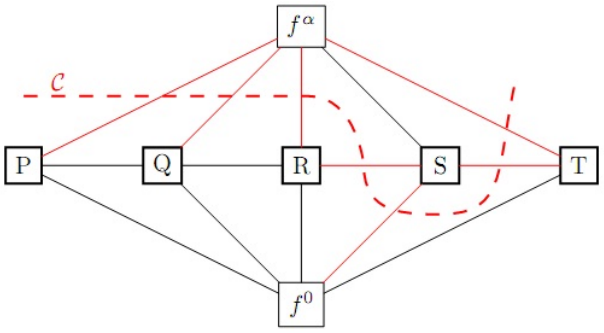

Figure 6. Cut through graph G using GC.

VI. APPLICATION TESTS

Two types of tests were conducted on developed application – calculation time test and depth map quality test.

Hardware configuration of computer where application was tested:

- CPU: Intel Core i5 3,20 GHz
- HDD: 500 GB 7200 RPM 16MB Cache
- RAM: 4 GB DDR 3 1066 MHz
- GPU: NVIDIA GeForce 275 GTX

A. Calculation Time Test

Calculation time test consisted from calculations of depth maps using BM, SGBM and GC algorithms. Input for this test was stereo image of simple scene (Fig. 7 shows left part of this input image) with several different image resolutions.

Resolutions of images which were used for creation of stereo images:

- 320 × 180
- 640 × 360
- 800 × 450
- 960 × 540
- 1280 × 720
- 1440 × 810
- 1920 × 1080

Depth map calculation times measured for individual algorithms and different image resolutions are shown in Table 1.

Figure 7. Left part of image used in Calculation Time Test.

TABLE 1. Calculation times

Resolution	BM (s)	SGBM(s)	GC(s)
320 × 180	0,080	0,127	7,363
640 × 360	0,227	1,137	48,707
800 × 450	0,366	1,438	130,225
960 × 540	0,471	1,803	220,660
1280 × 720	0,905	4,205	690,709
1440 × 810	1,179	9,628	800,940
1920 × 1080	2,882	11,378	3932,568

The fastest algorithm for dept map calculation is BM as can be seen in Table 1. The second one is SGBM and the last one is GC. Dependency between calculation time and image resolution for each one algorithm are also shown in Fig. 8.

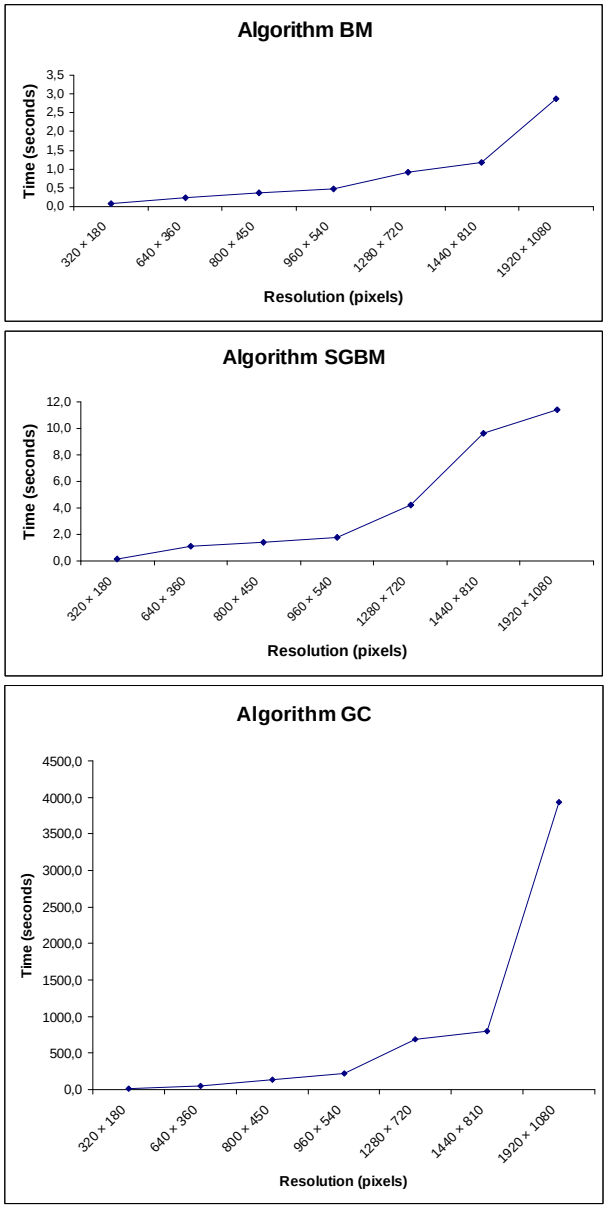

Figure 8. Dependency between calculation time and image resolution.

Comparison of individual algorithms in one graph is shown in Fig. 9. Subgraph for GC algorithm is scaled down 50 times, due to better fitting into graph.

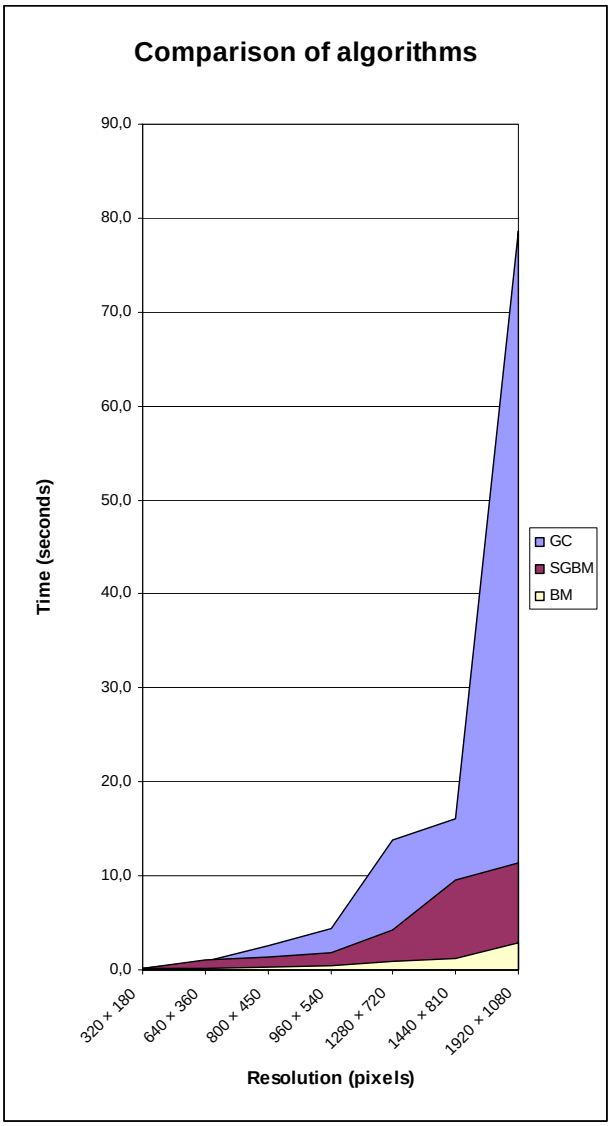

Figure 9. Comparison of algorithms.

Depth map calculation time is dependent from these six factors:

- image color,
- image resolution,
- number of objects in image,
- relative position of objects,
- objects distance,
- image background.

For example calculation time of depth map shown in Fig. 10 was 33 seconds. This depth map was created from input images with resolution 600 × 400 and with complex background Compare to this, calculation time of depth map shown in Fig. 11 was only 19 seconds and input images for this depth map has higher resolution (800 × 800) but has simpler scene with black background.

Figure 10. Example of image with lower resolution but with complex background.

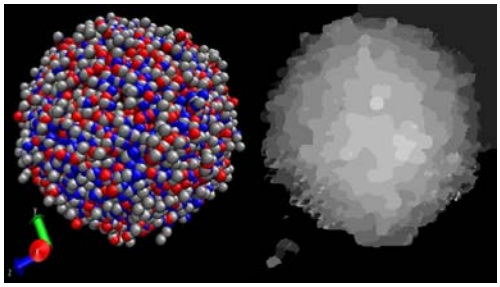

Figure 11. Example of image with higher resolution but with simpler scene and black background.

B. Depth Map Quality Test

This test shown, that the best quality has depth map calculated with GC, the second best quality has SGBM and the worst quality has BM. Left part of input image used in tests and depth maps calculated from this image by individual algorithms are shown in Fig. 12. – Fig. 15.

Figure 12. Input image (left part).

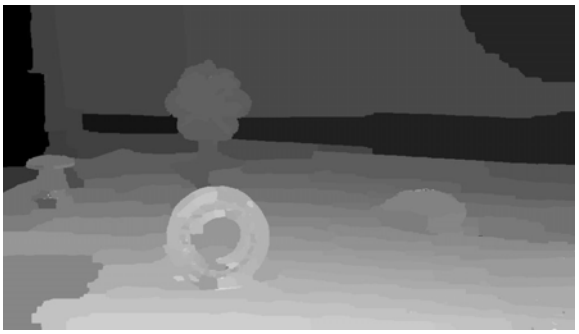

Figure 13. Depth map calculated with GC.

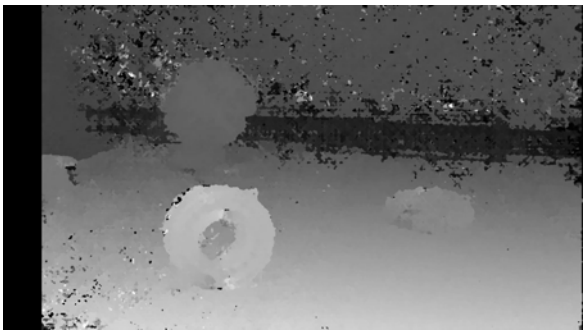

Figure 14. Depth map calculated with SGBM.

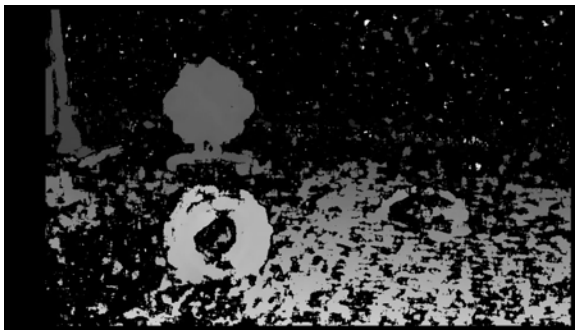

Figure 15 Depth map calculated with BM.

VII. CONCLUSION

The two most common passive autostereoscopic methods are parallax barrier and lenticular lens. Passive autostereoscopic 3D method from Philips uses another approach. For creation of 3D image is used 2D image and depth map – 2D-plus-depth method. This depth map can be created manually (which is very time consuming) or calculated from stereo image (video).

In this paper was presented application developed at DCI FEEI TU of Košice. This application calculates depth map from stereo image (video) using these algorithms: Block Matching, Semi-Global Block Matching and Graph Cut. Calculated depth map is than used as part of the input image for autostereoscopic 3D display Philips WOWvx (Fig. 16.).

Figure 16. Philips WOWvx with image created in tests.

Three algorithms are used for calculation: Block Matching, Semi-Global Block Matching and Graph Cut. Each one algorithm has its advantages and disadvantages. As can be seen from conducted tests (Fig. 8, Fig. 9, Fig. 12 and Table 1.), BM algorithm has fastest calculation time but worst quality of calculated depth map. GSBM offers better quality of depth map with also good calculation time. GC has longest calculation time but calculated depth map is very good.

Future work will be focused on improvement of calculation times for individual algorithms (especially GC). Work will be also focused on decreasing of calculation times for video which takes a lot of time. For example calculation time for one-minute video (25 fps × 60 seconds = 1500 images) with avarage calculation time of one image 0,34 second take more than eight hours. Solution to this can be usage of parallel computing environment based on cluster or GPGPU technologies.

ACNOWLEDGEMENT

This work is supported by VEGA grant project No. 1/0646/09: Tasks solution for large graphical data processing in the environment of parallel, distributed and network computer systems.

REFERENCES

[1] Sobota, B., Szabó Cs., Perháč, J., Ádám N.: 3D Visualization for City Information System, Proceedings of International Conference on Applied Electrical Engineering and Informatics AEI2008, Athens, Greece, 8.-11.9.2008, Košice, FEI TU Košice, 1, ISBN 978-80-553-0066-5, 9-13

[2] Xu, S.; Manders, C.M.; Odelia, T.Y.; Song, P.: 3D display for a classroom, Educational and Information Technology (ICEIT), 2010 International Conference on , vol.2, pp. V2-316-V2-320, 17. – 19. September 2010

[3] Philipse WOWvx homepage, http://www.business-sites.philips.com/3dsolutions/home/index.page

[4] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger, “Depth Map Creation and Image Based Rendering for Advanced 3DTV Services Providing Interoperability and Scalability”, Signal Processing: Image Communication. Special Issue on 3DTV, February 2007

[5] Boykov, Y.Y.; Jolly, M.-P.: "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," In ICCV 2001: Proceedings of the Eighth IEEE International Conference on Computer Vision, pp.105-112 vol.1, 2001

[6] Kim, J., Kolmogorov, V., Zabih, R.: Visual correspondence using energy minimization and mutual information, In ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision, pp. 1033 – 1040 vol. 2, Washington, DC, USA, 2003. IEEE Computer Society

[7] Ye Hou; Bao-long Guo; Jeng-Shyang Pan: "The Application and Study of Graph Cut in Motion Segmentation," Information Assurance and Security, 2009. IAS '09. Fifth International Conference on , vol.1, no., pp. 265 - 268, 18. – 20. August 2009

[8] Hui Wang; Hong Zhang; , "Adaptive shape prior in graph cut segmentation," Image Processing (ICIP), 2010 17th IEEE International Conference on , vol., no., pp.3029-3032, 26. – 29. September 2010

An Overview of Hierarchical Schemes for Fault Management in Wireless Sensor Networks

KHAN Muhammad Zahid[1], ASIM Muhammad[2], KHAN Ijaz Muhammad[3]
[1, 2] School of Computing and Mathematical Sciences,
Liverpool John Moores University, Byrom Street,
Byrom Street, L3 3AF, Liverpool, UK, E-Mail: M.Zahid-Khan@2008.ljmu.ac.uk[1], M.Asim@ljmu.ac.uk[2]

[3]Department of Computer Science,
Dhofar University,
211, Salalah, Oman, E-Mail: ijaz@edu.com

Abstract *— Wireless Sensor Networks (WSNs) are typically resource-constrained and battery driven. They are usually deployed in hostile and inaccessible environments to perform monitoring and tracking. Therefore, due to their physical deployment location sensor nodes are very prone to faults and failures. In recent years, much work has been done on the various aspects of the WSNs, especially fault management. Fault management is concerned with detecting, diagnosing, isolating and resolving faults and failures. Thereby, a network's management system with an efficient fault management platform makes the network fault tolerant in the events of faults and failures. In this context, many solutions have been proposed for fault management in WSNs. However, hierarchical architecture based schemes have proven to be more efficient as compare to centralized and distributed schemes. This paper aims to overview hierarchical architecture based schemes for fault management in WSNs. We critically analyze their effectiveness and short comings for large-scale WSNs. We believe through such an exercise provides a great background to establish new and effective fault management solutions for WSNs.*

Keywords**:** *WSNs, sensor networks, fault management, fault-tolerance, hierarchical architecture*

I. INTRODUCTION

WSNs are composed of a large number of self-organized sensor devices (homogenous and heterogeneous) that work in collaboration to monitor the physical environment and object of interest and relay messages to the Sink or Base Station. Sensor devices usually consist of a number of physical sensors, gathering environmental data like temperature or light, a microcontroller processing the data, and a radio interface to communicate with other nodes [1, 2]. These sensors have strict resource constraints and normally operate on batteries. WSNs have a variety of applications, such as military surveillance, industry monitoring, mass vehicle control and smart home, etc. [3].

These devices are typically resource-constrained and battery driven to allow autonomous work and wireless deployment in inaccessible terrain and hostile environments. Due to their physical deployment locations sensor nodes are expected to operate autonomously for a long period of time and may not be easily approachable for battery replacement and maintenance. Furthermore, harsh physical environment, e.g. rain, fire and falling of hard objects on senor hardware can also completely damage the device, hence faults and failures are normal facts WSNs [4, 5].

Thus, in order to guarantee the network quality of service and performance, it is essential for a WSN to be able to detect faults, and to perform something akin to healing and recovering from events that might cause faults or misbehaviour in the network, therefore fault management should be seriously considered in many wireless sensor network applications [5]. Fault management platform is an integral part of a network management system. Thereby, a network's management system with an efficient fault management platform makes the network fault tolerant and reliable in the events of faults and failures.

In this context many solutions have been proposed for fault management in WSNs. However, hierarchical architecture based schemes have proved to be more efficient as compare to centralized and distributed schemes. In this paper, we overview some of the most dominant hierarchical architecture based approaches developed for fault management in WSNs. Hierarchical based management is an efficient way of deploying fault management for large-scale WSNs. We critically analyze the effectiveness and short comings of these schemes for WSNs. We believe through such exercise we provide a great background to establish new and effective fault management solutions for WSNs.

The rest of the paper is organized as follows. Ssection II, gives a background about faults and fault management in WSNs. In section III dominant fault management schemes based on hierarchical architecture have been critically analysed. It followed by discussion in section IV and conclusions and future work is given in section V.

II. BACKGROUND

To comprehend fault management, it is important to point out the difference between faults, and failures. A fault is any kind of defect that leads to an error. A failure is a state which occurs when the system deviates from its specification and cannot deliver its intended functionality. Liu et al. [6], classify fault tolerance into four levels from

the system point of view such as: hardware layer, software layer, network communication layer, and applications layer. In fault management research literature [7, 8], node hardware fault has been categorized into four types such as: *Permanent faults*, *Intermittent faults*, *Temporary faults* and *Potential faults*. Permanent faults are continuous and stable in nature, e.g. hardware faults within a component. An intermittent fault has the occasional (such as a regular or irregular interval) manifestation due to unstable characteristics of the hardware. These faults are the result of some temporary environmental impact on otherwise correct hardware, e.g. the impact of cosmic radiation on the sensor [7]. Potential faults usually occur due to the depletion of node hardware resources, such as node's battery energy exhaustion. Such faults can cause node's sudden death.

A. Fault Management

Fault management is a very important component of network management concerned with detecting, diagnosing, isolating and resolving faults and errors. Fault management can be defined as a set of services and functions performed to detect, diagnose, isolate and rectify malfunctions in a network. It also involves compensation for environmental changes, monitoring and examining errors' logs, accepting and acting on error detection, tracing and identifying faults. Furthermore, carrying out sequences of diagnostic's tests, correcting faults and failures, reporting error conditions and localizing and tracing faults are part of the fault management functions [9]. Important functions of fault-management include:

- Definition of thresholds for potential failure conditions
- Constant monitoring of system status and usage level
- General diagnostics
- Alarm and the notification of any error or malfunctions
- Tracing the location of potential and actual malfunctions
- Automatic correction of potential-problem causing conditions
- Should keep the probability of false alarm as minimum as possible
- Recovery of failures

Fault management for WSNs is different from traditional networks. Recent research has developed several schemes and techniques that deal with different types of faults at different layers of the network. Fault management in WSNs can be classified according to their management system network architecture [10, 11]: *Centralized, Distributed, or Hierarchical*. In a centralized management architecture, the base station acts as a central controller or a central manager station that collects information from the whole network and controls the entire network. Instead of having a single central controller, distributed management architecture employs multiple manager station throughout the whole network. Each manager controls a sub-region of the network. Hierarchical management architecture is a hybrid between the centralized and distributed approach. Sub-controller or managers are distributed throughout the network in a tree shape hierarchical manner, having levels of lower and higher level of hierarchy.

To provide resilience in faulty situations three main actions (fault detection, fault diagnosis and fault recovery) (Figure - 1) must be performed [8, 12-14].

1) **Fault Detection -** Fault detection is the first phase of fault management, where an unexpected failure in the network should be properly identified by the networks system. Fault detection in sensor networks largely depends on the type of applications and the type of failures.
2) **Fault Diagnosis -** Fault diagnosis is a stage in which the causes of detected faults can be properly identified and distinguished from other irrelevant alarms.
3) **Fault Recovery -** After fault detection and fault diagnosis; it is seen in fault recovery that how faults can be treated [13]. The failure recovery phase is the stage at which the sensor network is restructured or reconfigured, in such a way that failures or faulty nodes do not impact further on network performance [8].

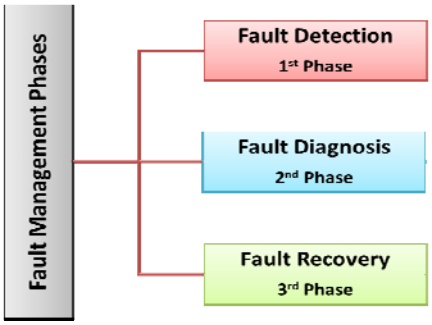

Figure 1. Fault Management Phases.

III. HIERARCHICAL ARCHITECTURE BASED SCHEMES FOR FAULT MANAGEMENT

Hierarchical management architecture is a hybrid between the centralized and distributed approaches. Sub-controller or managers are distributed throughout the network in a tree shape hierarchical manner, having levels of lower and higher level of hierarchy. Hierarchical model distributes fault management tasks according to the node management functionalities and responsibilities in the network. It splits the whole network into several regions. Each region consists of a limited number of sensor nodes. A manager node is selected to be responsible for the fault management within its region [7].

It is also important to have continuous connectivity in a WSN after its deployment in a hostile environment [15]. Clustering has become an emerging technology for building hierarchical and scalable applications for WSNs. Most of the contemporary fault management architectures have used the clustering-based hierarchical approaches.

An important objective of any clustering technique is to maintain network connectivity and balance the energy consumption for resource constrained WSNs. Mention LEACH [16], HEED,[17] [19] GAF [18]. These are some prominent clustered-based approaches that employ the clustering mechanism that proved to make the network fault tolerant and reliable, hence to prolong the network lifetime.

In clustering paradigm, sensor nodes in the network are grouped together to efficiently relay the sensed data to the Sink/Base Station. Each group of sensor or cluster nodes has a cluster head and gateway node. In a cluster based sensor networks, when sensors are first activated, the neighboring sensor nodes organize themselves into clusters to reduce the sensing redundancy and to avoid the reuse of scarce limited resources.

Examples are: Distributed fault detection by using clustering mechanisms [19-21] and LEACH [22]. LEACH (Low Energy Adaptive Clustering Hierarchy) is the most famous protocol. It is a self-organizing, adaptive clustering protocol that selects cluster heads randomly to distribute the energy load evenly among the sensor nodes in the network. The role of a cluster head is rotated randomly in order to prevent the energy drainage of a particular single sensor node.

However, the randomized rotation does not take into account the current energy level of the node and may choose a node with very little remaining energy as a cluster head with the danger of fast death of that node. Furthermore, this algorithm allows only 1-hop clusters to be formed that may lead to the formation of a large number of clusters.

The approach causes a problem with energy-efficiency and scalability, because when the network size grows the cluster head will not be possible to reach the Sink or base station [23]. LEACH-C (LEACH-Centralized) [23] is an improved version of LEACH, which forms clusters at the beginning of each round using a centralized decision making algorithm. LEACH-C selects cluster heads based on their location information and energy level. LEACH-C performs well, but frequent communication between the base station and sensor nodes increase communication cost and energy usage.

Hierarchical Clustering introduces an extra level of management nodes that facilitate the distribution of control over the entire network. It saves energy and reduces network contention by enabling locality of communication: nodes communicate their data to their cluster head over a short distance, while these cluster heads further forward data to their high-level manager in the hierarchy or directs it to the base station [24].

Most of the existing hierarchical clustering approaches assume a single hop communication model in terms of members' nodes. For instance, Siqueia et al. [25], proposed a 3-tier hierarchical clustering architecture, which has a single hope communication model between cluster-head and sensor nodes, or between cluster-heads and the base station. They work well for small networks, but their performance is heavily impaired when the number of clusters increased in large-scale sensor networks.

Gupta *et al.* proposed a fault-tolerant clustering architecture for WSNs [26, 27], where the gateway node constructs the overall network communication map by sharing and exchanging their own local information with other gateway nodes, and if the node does not get information from another node then it is clear that the node is faulty (See Figure - 2). When a gateway node dies, the cluster is dissolved and all its nodes are reallocated to other healthy gateways. The re-clustering process consumes more time and energy; in addition, all the cluster members are involved in the recovery process, which minimize the overall network lifetime.

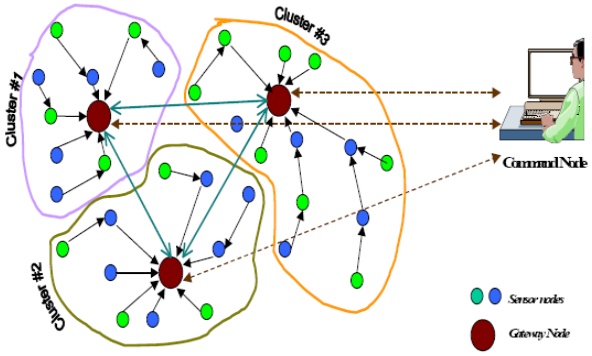

Figure 2. Fault-Tolerant Clustering for WSNs.

Localized Cluster-Based method for fault detection and network connectivity recovery, which is energy-efficient and responsive is proposed by Venkataraman *et al.* in [20]. The scheme considers only permanent faults, which occur mainly due to energy depletion, in particular, and, which leads to the loss of connectivity and coverage in the network. It uses a cluster approach to organize the nodes in a tree-like manner with a parent and children nodes. In this particular type of clustering mechanism, failure is detected by the *fail_report_msg* disseminate by the failure node to its parent and child (one hop neighbour), when its energy level drops below the threshold value *Eth*.

This information of the failure report is an indication to start the failure recovery process, and the corrective or recovery process is taken only by those nodes that have the information. Therefore, energy is saved by not allowing all the nodes in the cluster to detect a failure. Further, clustering approach helps in connectivity maintenance by re-organizing clusters, saves energy, avoid contention by enabling locality of communication. The localized fault detection method has been found energy-efficient in comparison with another algorithm proposed by Chessa, et.al. [28] and the network connectivity recovery method is more efficient than Gupta [26] algorithm and the Crash fault identification method [29]. From the simulated results, it is clear that the proposed algorithm is quicker due to the localized decision making using the clusters. The method consumes less energy since no additional replacement or movement is required to recover the fault. However, the main drawback is that it can only detect permanent faults and avoids intermittent and transient faults.

To improve the robustness and efficiency of clustered-based scenario, Lai and Chen [30] proposed a CMATO (Cluster-Member-based fAult Tolerant mechanism) algorithm. CMATO views the cluster as a whole and takes advantage of the inter-cluster monitoring of nodes to detect the faults. When the cluster member detects a fault that is caused by the cluster head, they act co-operatively to select new cluster head to replace the failed one. The CMATO fault-tolerant mechanism works on the idea of overhearing of nodes in of their cluster-head and other members, so that the node can detect the failure of the cluster head quickly. When all the nodes aware the failure some nodes join the neighboring clusters, while others join the newly constructed cluster. However, this method would increase the number of cluster heads, and that would cause the interference among clusters when multiple medium faults happen in the network.

WSNMP (Wireless Sensor Networks Management Protocol) [31] is a hierarchical network management system which also uses tier architecture. In this approach, a central manager is set at the highest level of the network i.e. the sink node; the intermediate manager works at the cluster heads and management agents are the normal sensor nodes. Intermediate managers are used to distribute management functions and collect and collaborate with management data from the entire network. They execute management functions based on their local network states whereas central network manager has the global knowledge of the network states and entire topology map. Once the topology of the network is modeled the Central Manager can reconfigure the network with minimum overhead. It also detects a fault in the network by identifying the non-response nodes and if required to reconfigure the routing path [31]. The architecture of WSNMP is shown in figure-3 [31], which represents the relationship between management services and management functionalities. WSNMP provides the method to monitor the network states by collecting management data and accordingly control and maintain the network resources. However, to construct the entire topology map for the whole network incurs extra over-head and is more energy consuming for resource constrained wireless sensor networks.

Asim et al. presented a new cellular based fault management architecture to achieve distributed fault detection and fault recovery. The scheme divides the network into a virtual grid/cell where each cell consists of a group of nodes, which supports scalability and performs fault detection recovery locally with minimum energy consumption. The grid-based architecture permits the implementation of fault detection in a distributed manner and allows the failure report to the forwarded across cells. A cell manager and a gateway node are chosen in each cell to perform management tasks locally. A hierarchy of different nodes is formed, and each node is assigned a specific role in the hierarchy such as Cell Manager, Group Manager, etc. Cell manager and gateway nodes coordinate with each other to detect faults with minimum energy consumption. The architecture assumes a homogenous network where all nodes are equal in resources and can easily back up each other in case of recovery. However, it only considers permanent faults and there is no remedy for Intermittent and transient faults.

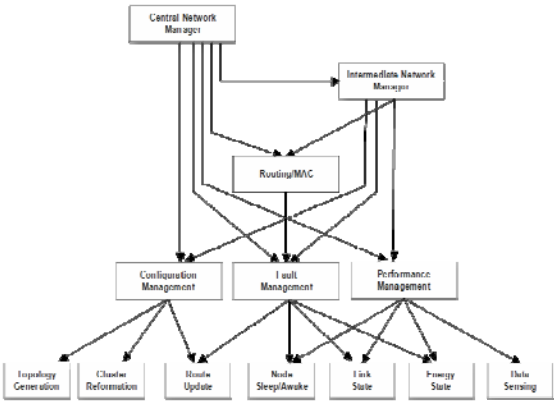

Figure 3. WSNMP Fault Management Architecture.

Yu et al., proposed [7] a new hierarchical fault management architecture for WSNs. The authors classify the sensor nodes into four management roles such as: Common nodes, Managing nodes, Group manager, and Base station. The paper proposed a concept model in a tree structure to describe faults monitored in sensor networks. The “Nodes” describes the fault occurred in a node, i.e. potential failure and the “Nodes” describes the network faults caused by either the potential or permanent failure of one or a set of sensor nodes. The proposed fault management architecture adopts two types of fault detection modes such as: the node self-detection and passive detection to enhance the accuracy and efficiency of fault management. Self-detection is used to monitor the residual hardware status periodically and to identify the potential faults. To efficiently detect the node sudden death management components adopt a passive detection mode where Group managers and Managing nodes take most parts of passive detection to monitor sensor nodes in the network. The authors evaluate the approach in the GTSNets network simulator and analyze its costs by measuring node energy expenditure. The experiments demonstrate that the self-manageable WSNs may be more energy-efficient than a centralized approach, especially in long sensor network connectivity nodes in a multi-hop communication. However, it fails to detect a faulty node if a node is disconnected from the network due to some reason such as link failure.

Hierarchical clustering based distributed approaches provide a major shift in the design of fault management architecture for WSNs. Management responsibilities are transferred more towards the sensor nodes, instead of a central manager, which ultimately makes the network more reliable and self-managed. However, most of the schemes discussed earlier, are not fully adaptive and self-managed. Fault management operations such as fault detection, diagnosis and recovery are carried out by exchanging excessive messages between nodes. To overcome this problem Yu et al. [7, 32], proposed a

biologically inspired self-managed fault management architecture for WSNs. Self-managed fault management means that a WSN must perform fault management tasks and services with a minimum or no human intervention with the goal of promoting network productivity and quality of service [33]. The self-managed fault tolerant WSNs must be able to detect and recover from various networks and sensor faults locally in a distributed way with minimum resource utilization [34]. The proposed self-managed hierarchical architecture fully distributes the management tasks among different sensor nodes in the network. The paper, particularly tries to examine the self-management capabilities adapting to various requirements (e.g. sensor node failure) in a rapidly changing and hostile environment. Instead of considering the stereotype distributed clustering technique, the authors introduce a new management layer between the cluster-head and its leaf nodes. The scheme introduces more self-managing functions to the sensor nodes, which encourages them to be more self-dependent on monitoring their own status instead of frequent consulting with their cluster-head. In additions, they also give a solution for faulty node's replacement in a self-configurable WSN.

IV. Discussion

In the previous section, we discussed and analyzed some of the most dominant schemes in fault management based on hierarchical architecture. The use of a hierarchical arrangement has various intrinsic advantages. Sensor nodes, which are normally unable to communicate due to limited radio signals, can be interconnected. Furthermore, dividing large WSNs into smaller sub-networks decreases the number of hops required to reach each sensor node that results in better communication performance with lower delay and less packet loss. Another notable advantage of the hierarchal structure is that heterogeneous sensor node platform can be easily integrated into the heterogeneous WSN. The hierarchical structure manages large scale WSNS efficiently by distributing the management load from a single entity to various managing nodes in the hierarchy. More management tasks are assigned to the nodes which less energy-constrained, and the resource-constrained sensor nodes have fewer management functions to perform. This careful distribution of management tasks will highly reduce the energy, memory and computational requirements.

Hierarchical clustering introduces an extra level of management nodes that facilitate the distribution of control over the entire network. It saves energy and reduces network contention by enabling locality of communication: nodes communicate their data to their cluster head over a short distance, while these cluster heads further forward data to their high-level manager in the hierarchy or directs it to the base station [24]. However, most of the hierarchical clustered based schemes do not take into account the number of cluster head that is created. Because the excessive number of cluster heads would cause the interference among clusters. Furthermore, the location of the cluster-head and the number of sensor is each cluster is also an important issue to be considered.

We, therefore, contend that there is still a need of a new fault management scheme to address all the problems in existing fault management approaches for WSNs. We must take into account a wide variety of sensor applications with diverse needs, different sources of faults, and with various network configurations. In addition, it is also important to consider other factors, i.e. mobility, scalability and timeliness.

V. Conclusion and Future Work

Fault management has been widely considered as a key part of today's network management, especially in the context of WSNs. Recent rapid growth of interests in WSNs has further strengthened the importance of fault management, or in particular, played a crucial role. The contribution of this paper is to present an in-depth critical overview of some of the most dominant hierarchical architecture based schemes for fault management in WSNs. Hierarchical model distributes fault management tasks according to the node management functionalities and responsibilities in the network. Hierarchical clustered based architecture introduces an extra level of management nodes that facilitate the distribution of control over the entire network. It saves energy and reduces network contention by enabling locality of communication. In future, we will be conducting simulation experiments to evaluate the performances of the discussed schemes, both on their own and against schemes of similar types. We are committed to share our research findings with the ongoing research in this area.

Acknowledgment

The author acknowledges University of Malakand (UOM), KPK, Pakistan, and Higher Education Commission (HEC) of Pakistan for funding towards his PhD project.

References

[1] L. M. d. Souza, H. Vogt, and M. Beigl, "A survey on Fault Tolerance in Wireless Sensor Networks," *[Online]. Available: http//:digbib.ubka.uni-karlsruhe.de/volltexte/documents/11824*, 2007.

[2] M. Z. Khan and I. M. Khan, "A Research Based Review of Wireless Sensor Networks," *Annals. Computer Science Series*, vol. 9, p. 16, Dec.2011.

[3] W. Dargie and C. Poellabauer, *FUNDAMENTALS OF WIRELESS SENSOR NETWORKS THEORY AND PRACTICE*: A John Wiley and Sons, Ltd., Publication, 2010.

[4] M. Z. Khan, M. Merabti, and B. Askwith, "Design Considerations for Fault Management in Wireless Sensor Networks," in *Proceedings of the 10th Annual PostGradute Symposium on The Convergence of Telecommunications, Networking and Broadcasting, PGNet 2009, Liverpool John Moores University*, Liverpool, UK, June 2009, pp. 3-9.

[5] M. Z. Khan, M. Merabti, B. Askwith, and F. Bouhafs, "A Fault-Tolerant Network Management Architecture for Wireless Sensor Networks," presented at the 11th Annual PostGradute Symposium on The Convergence of Telecommunications, Networking and Broadcasting, PGNet 2010, Liverpool John Moores University, Liverpool, UK, June 2010.

[6] H. Liu, A. Nayak, and I. Stojmenovic, "Fault-Tolerant Algorithms/Protocols in Wireless Sensor Networks," in *Guide to Wireless Ad Hoc Networks*, ed: Springer-Verlag London, 2009, pp. 265-295.

[7] M. Yu, H. Mokhtar, and M. Merabti, "Self-Managed Fault Management in Wireless Sensor Networks," in *Proceedings of the The Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM '08)*, 2008, pp. 13-18.

[8] Y. Mengjie, H. Mokhtar, and M. Merabti, "Fault Management in Wireless Sensor Networks," *IEEE Wireless Communications*, vol. 14, pp. 13-19, 2007.

[9] M. Al-Kasassbeh and M. Adda, "Network fault detection with Wiener filter-based agent," *Journal of Network and Computer Applications*, vol. 32, pp. 824-833, 2009.

[10] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A Survey on Sensor Networks," *IEEE Communication Magazine*, pp. 102-114, 2002.

[11] W. L. Lee, A. Datta, and R. Cardell-Oliver, *Network Management in Wireless Sensor Networks*: Handbook on Mobile Ad Hoc and Pervasive Communications American Scientific Publishers, 2006.

[12] L. M. d. souza, H. Vogt, and M. Beigl, "A survey on Fault Tolerance in Wireless Sensor Networks," *www.digbib.ubka.uni-karlsruhe.de/volltexte/documents/11824.* , n.d.

[13] L. Paradis and Q. Han, "A Survey of Fault Management in Wireless Sensor Networks," *Journal of Network and System Management, Springer Science + Business Media, LLC*, vol. 15, pp. 171-190, June 2007.

[14] M. Asim, Hala.Mokhtar, M. Z. Khan, and M. Merabti, "A Sensor Relocation Scheme for Wireless Sensor Networks," presented at the Fifth International Workshop on Telecommunication Networking, Applications and Systems (TeNAS 2011), Workshops of IEEE 25th International Conference on Advanced Information Networking and Applications (AINA 2011), Biopolis, Singapore, March 2011.

[15] O. Younis, M. Krunz, and S. Ramasubramanian, "Node clustering in wireless sensor networks: recent developments and deployment challenges," *IEEE Networks*, vol. 20, pp. 20-25, 2006.

[16] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, pp. 660-670, 2002.

[17] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, pp. 366-379, 2004.

[18] W. Dali, H. A. Chan, and K. V. N. Kameri, "Circular-Layer Algorithm for Ad Hoc Sensor Networks to Balance Power Consumption," in *3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks, SECON '06*, 2006, pp. 945-950.

[19] A. T. Tai, K. S. Tso, and W. H. Sanders, "Cluster-based failure detection service for large-scale ad hoc wireless network applications," in *Proceedings of the International Conference on Dependable Systems and Networks*, 2004, pp. 805-814.

[20] G. Venkataraman, S. Emmanuel, and S. Thambipillai, "A Cluster-Based Approach to Fault Detection and Recovery in Wireless Sensor Networks," presented at the 4th International Symposium on Wireless Communication Systems, ISWCS'07, 2007.

[21] C. Yao-Chung, L. Zhi-Sheng, and C. Jiann-Liang, "Cluster based self-organization management protocols for wireless sensor networks," *IEEE Transactions on Consumer Electronics*, vol. 52, pp. 75-80, 2006.

[22] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," presented at the 33rd Annual Hawaii International Conference on System Sciences, 2000.

[23] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer Communications*, vol. 30, pp. 2826-2841, 2007.

[24] V. Srikanth and I. R. Babu, "Cluster Head Selection for Wireless Sensor Networks: A Survey," *The Icfai University Journal of Information Technology*, vol. 5, pp. 44-53, March 2009.

[25] I. G. Siqueira, L. B. Ruiz, A. A. F. Loureiro, and J. M. Nogueira, "Coverage area management for wireless sensor networks," *Int. J. Netw. Manag.*, vol. 17, pp. 17-31, 2007.

[26] G. Gupta and M. Younis, "Fault-Tolerant Clustering of Wireless Sensor Networks," in *IEEE Wireless Communications and Networking, WCNC'03*, 2003, pp. 1579-1584.

[27] G. Gupta and M. Younis, "Load-Balanced Clustering in Wireless Sensor Networks," presented at the Proceedings of International Conference on Communication (ICC 2003), Anchorage, AK, 2003.

[28] S. Chessa and P. Maestrini, "Fault recovery mechanism in single-hop sensor networks," *Computer Communications*, vol. 28, pp. 1877-1886, 2005.

[29] S. Chessa and P. Santi, "Crash Faults Identification in Wireless Sensor Networks," *Computer Communications*, vol. 25, pp. 1273-1282, 2002.

[30] L. Yongxuan and C. Hong, "Energy-Efficient Fault-Tolerant Mechanism for Clustered Wireless Sensor Networks," in *Proceedings of 16th International Conference on Computer Communications and Networks (ICCCN'07)*, 2007, pp. 272-277.

[31] M. M. Alam, M. Mamun-Or-Rashid, and C. S. Hong, "WSNMP: A Network Management Protocol for Wireless Sensor Networks," presented at the 10th International Conference on Advanced Communication Technology, (ICACT'08) 2008.

[32] M. Yu, H. Mokhtar, and M. Merabti, "A self-organised middleware architecture for Wireless Sensor Network management," *International Journal Ad Hoc Ubiquitous Comput.*, vol. 3, pp. 135-145, 2008.

[33] L. B. Ruiz, *et al.*, "On the design of a self-managed wireless sensor network," *IEEE Communications Magazine*, vol. 43, pp. 95-102, 2005.

[34] W. L. Lee, A. Datta, and R. Cardell-Oliver, "WinMS: Wireless Sensor Network-Management System, An Adaptive Policy-Based Management for Wireless Sensor Networks," *School of Computer Science & Software Engineering, The University of Western Australia, CSSE Technical Report UWA-CSSE-06-001*, June 2006.

Optimization Solution for Multiple Model Control Structures

LUPU Ciprian[1], PETRESCU Catalin[2]

[1] University POLITEHNICA of Bucharest, Romania,
Department of Automatic Control and Systems Engineering, Faculty of Automatic Control and Computers,
313, Splaiul Independentei, 060042, Bucharest, Romania, cip@indinf.pub.ro

[2] University POLITEHNICA of Bucharest, Romania,
Department of Automatic Control and Systems Engineering, Faculty of Automatic Control and Computers,
313, Splaiul Independentei, 060042, Bucharest, Romania, catalin@indinf.pub.ro

Abstract – ***The multiple model structures are specific control solutions for some classes of systems with important nonlinearities or different functioning regimes. One of these structures' specific problems is the determination of the models number: an increased number leads to superior performances but very complex structure. The paper presents an original methodology for models number reducing without decreasing the performances.***
This solution is of practical importance allowing facile implementation on PLC and process computers. The experimental results prove the structure's performances.

Keywords: ***multiple models, nonlinear process, compensator, structure optimization.***

I. INTRODUCTION

The multi model systems represent a relative new approach for the nonlinear systems control. Since the 90's, different studies on multi model control strategies have been developed. The Balakrishnan's and Narenda's first papers propose several stabile and robust methods using classical switching and tuning algorithms [1].

Further research in this field determined the extension and improvement of the multi model control concept. Magill and Lainiotis introduce the model representation through Kalman filters. In order to maintain the stability of the minimum phase systems, Middelton improves the switching procedure using an algorithm with hysteresis. Landau and Karimi have important contributions regarding several particular multiple model adaptive structures [2]. Dubois, Dieulot and Borne apply fuzzy procedures for switching and use sliding mode control.

This paper proposes a multi model control structure which contains, for each model/controller pair, a nonlinearity compensator. It is based on the determination of each model's static characteristic. This solution reduces the number of models and a decreases the overall complexity of the global structure.

This structure can be applied in the case of processes with important nonlinearities.

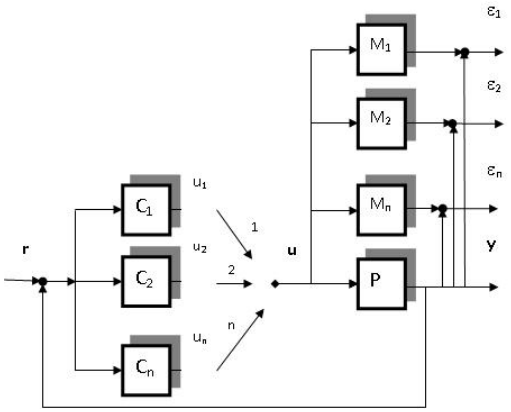

Figure 1. Multi model control structure.

II. CLASSIC MULTI MODEL APPROACS

The classic control solution implies choosing a set of models *M*, and on a set of the correspondent controllers *C*:

$$M = \{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3 \ldots \mathcal{M}_n\},$$
$$C = \{C_1, C_2, C_3 \ldots C_n\} \quad (1)$$

Based on these model/controller pairs the closed-loop configuration is the one presented in Fig. 1.

The input and output of the process *P* are *u* and *y* respectively, and *r* is the set point of the system. The M_i ($i=1, 2, \ldots n$) models are determined a priori. For each model M_i a controller C_i is designed in order to assure the nominal performances for the pair (M_i, C_i).

The main idea of the multi model structure construction is based on dividing the process functioning region in *n* small disjoint and adjacent zones, for which the models are simpler and the *n* corresponding control algorithms have low complexity (Fig. 2).

One of the principles used in zones' choosing is that the absolute value of the difference between the static characteristic and its linearization has to be smaller than the imposed threshold.

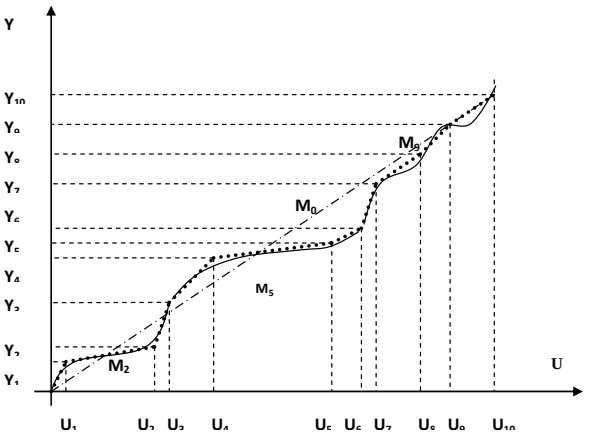

Figure 2. Construction of the set of process's models.

This does not impose using a linear model for the region. It is very possible to have a second or third or *m* order model and a complex corresponding control algorithm. A very complex algorithm can determine better performances but uses important hardware resources on real time implementation.

In real situations there must be a balance between complex control algorithms and complex real time hardware/software architectures.

In Fig. 2 the continuous line represents the process static characteristic, the dotted line the linear models for a large number of zones and dashed and dotted line is global linear model.

The difference between the global model and the process characteristic is large (maximal distance in *U6, Y6* point).

Using a single controller provides poor performances. For high performances and robust implementation a more complex control strategy must be used.

III. PROPOSED SOLUTION FOR MULTIPLE MODEL STRUCTURE

This paper proposes a multi model control structure which, for each controller, provides a nonlinearity compensator [3]. This solution allows a reduced number of models and a reduced complexity for each control algorithm. This solution is named "control system with inverse model" [10].

In literature proposes a lot of inverse model structures. For the presented control solution a very simple and efficient structure, presented in Fig. 3, is employed. This solution sums two commands: the first one "a direct command" generated by the feed forward command generator, and the second generated by a classic and very simple algorithm (PID, RST etc.).

This structure is added to all the model/controller pairs of the multi model structure. For each controller, the first command, based on the process static characteristics, is dependent on set point value and is designed to generate a corresponding value to drive the process's output close to imposed set point value. The second (classic feedback) algorithm generates a command that corrects the difference caused by external disturbances and, accordingly to the set point, by eventual bias error caused by mismatches between calculated inverse process characteristic and the situation from real process.

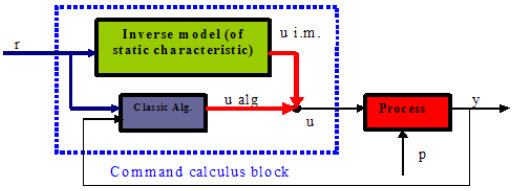

Figure 3. Proposed scheme for inverse model structure.

The presented solution proposes treating these "inverse model" mismatches that "disturb" the first command as a second command classic algorithm's model mismatches. This solution imposes designed a classic algorithm with robustness reserves. For this reason, designing the second algorithm has in two steps:

- design of a classic algorithm based on a model identified in a functioning point – selected fortuitously or, on the middle of the corresponding segment process characteristic;
- verification of algorithm's robustness and improving, if necessary - (re)designing procedure;

In Fig. 3, the blocks and variables are as follows: Process – physical system to be controlled; Command calculus – unit that computes the process control law; Classic Alg. – control algorithm (PID, RST); y – output of the process; u – output of the Command calculus block; u alg. – output of the classic algorithm; u i.m. – output of the inverse model block; r – system's set point or reference trajectory; p – disturbances of physical process.

This solution used in the context of a multi model structure has three important aspects:

- Selection of a reduced number of zones where the nonlinearity is important but lower than an imposed threshold.
- Construction of the compensator block for each zone.
- Designing the correspondent controller for each zone.

All three will be presented in next sections.

A. Zones Selection

The number of zones must be reduced (*2, 3* or maximum *4*) and these can consist in the medium or "local" tendencies of the nonlinear characteristic [4], [5]. Fig. 4 presents an example for this aspect.

It can be imposed that the difference between the tendency and the real characteristic must be less or equal to an imposed margin.

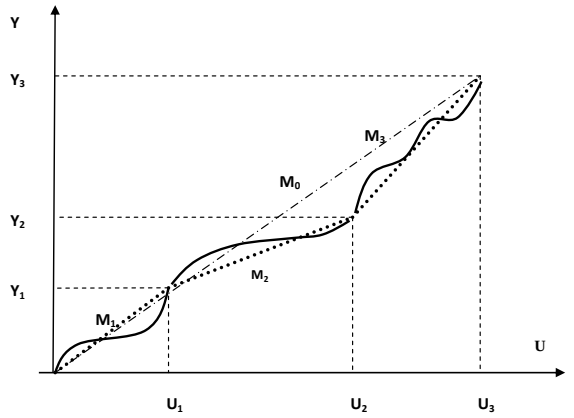

Figure 4. Selection of major zones.

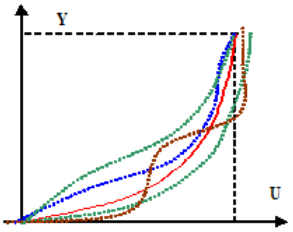

Figure 5. Determination of static characteristic of the process. Red (continuous) line represents the final characteristic.

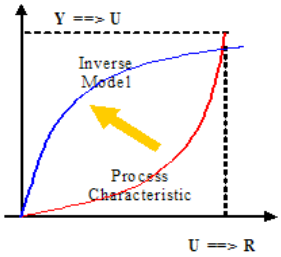

Figure 6. Construction of inverse model.

In Fig. 4 the continuous line represents the process static characteristic, the dotted line – the linear models and dashed and dotted line the global linear model.

B. Construction of Nonlinear Compensator Blocks

This operation is based on several experiments. The command *u(k)* is increasing and decreasing and the corresponding stabilized process output *y(k)* is measured. The command *u(k)* covers all possible values (*0 to 100*% in percentage representation). Because the process is disturbed by noises, the measurements of the static characteristics are not identically. The final static characteristic is obtained by meaning these experiments. Fig. 5 presents this operation. The graphic between two "mean" points can be obtained using an extrapolation procedure.

According to system identification theory [4] the dispersion of process trajectory can be found using expression (2):

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} \tag{2}$$

This can express a measure of noise, process's nonlinearity etc. and is important for robustness perspective during the design of the control algorithm [7], [8].

The next step in obtaining the nonlinear compensator block deals with inverting the process's static characteristic. Fig. 6 presents this construction. According to this, *u(k)* is dependent to *r(k)*. This characteristic is stored in a table; thus, for the inverse model based controller, selecting a new set point *r(k)* means searching in this table the corresponding command *u(k)* that determines a process output *y(k)* close to the reference value.

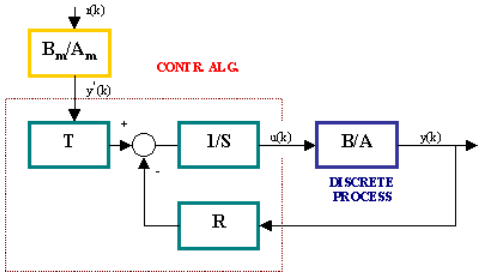

Figure 7. RST control algorithm structure.

C. Controllers Design

The zones control algorithm's duty is to eliminate the disturbances and differences between inverse models' computed command and real process behavior. A large variety of control algorithms can be used here, PID, RST, fuzzy etc., but the goal is to have a simplified one.

For this study we use a RST algorithm. This is designed using a pole placement procedure [2]. Fig. 7 presents the RST algorithm:

Where R, S, T polynoms are:

$$\begin{aligned} R(q^{-1}) &= r_0 + r_1 q^{-1} + \ldots + r_{nr} q^{-nr} \\ S(q^{-1}) &= s_0 + s_1 q^{-1} + \ldots + s_{ns} q^{-ns} \\ T(q^{-1}) &= t_0 + t_1 q^{-1} + \ldots + t_{nt} q^{-nt} \end{aligned} \tag{3}$$

Pole placement algorithm makes use of the identified model.

$$y(k) = \frac{q^{-d} B(q^{-1})}{A(q^{-1})} u(k) \tag{4}$$

where

$$\begin{aligned} B(q^{-1}) &= b_1 q^{-1} + b_2 q^{-2} + \ldots + b_{nb} q^{-nb} \\ A(q^{-1}) &= 1 + a_1 q^{-1} + \ldots + a_{na} q^{-na} \end{aligned} \tag{5}$$

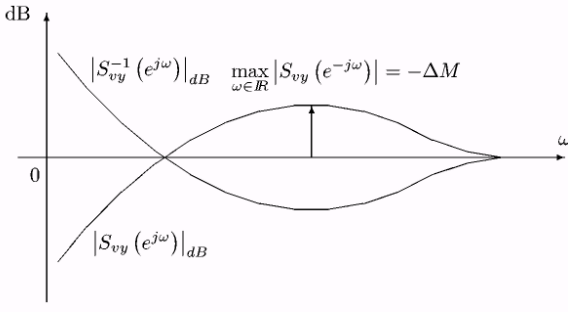

Figure 8. Sensitivity function graphic representation.

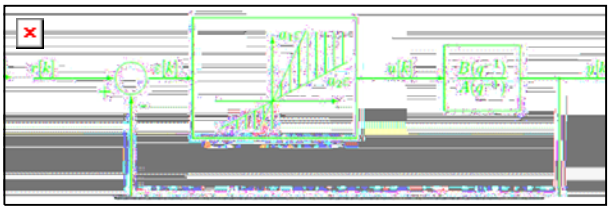

Figure 9. Robust control design procedure.

The identification is made in a specific process operating point and can use recursive least square algorithm developed in [2].

This approach allows the users to verify, and if is necessary, to calibrate the algorithm's robustness. The following expression and Fig. 8 present the "disturbance-output" sensitivity function.

$$S_{vy}(e^{j\omega}) \overset{def}{=} H_{vy}(e^{j\omega}) = \\ = \frac{A(e^{j\omega})S(e^{j\omega})}{A(e^{j\omega})S(e^{j\omega}) + B(e^{j\omega})R(e^{j\omega})}, \quad \forall \omega \in R \tag{6}$$

The negative maximum value of sensitivity function represents the module margin.

$$\Delta M\big|_{dB} = -\max_{\omega \in R} \left| S_{vy}(e^{j\omega}) \right|_{dB} \tag{7}$$

Based on this value [2], in a "input-output" representation, process nonlinearity can be bounded inside of "conic" sector, presented in Fig. 9, where a_1 and a_2 are calculated using next expression:

$$\frac{1}{1-\Delta M} \geq a_1 \geq a_2 \geq \frac{1}{1+\Delta M} \tag{8}$$

Finally, if it is imposed that all nonlinear characteristics should be (graphically) bounded by the two gains, or if the gain limit should be greater or equal to process static characteristic maximal distance $\Delta G \geq md$, then a controller that has sufficient robustness was designed.

D. Global Architectures

Partitioning the nonlinear characteristic like in Fig. 4 and combining the multi model structure (presented in Fig. 1) with the control structure (presented in Fig. 4 determines) the global architecture of multi model control system presented in Fig. 10.

On Fig. 10, the blocks and variables are as follows: Process – physical system to be controlled: Command calculus – unit that computes the process control law; Alg. *i* – i control algorithms (PID, RST); *y* – output of the process; *u* – output of the Command calculus block; *u i.* – output of the *i* control algorithm; *r* – system's set point or reference trajectory; *p* – disturbance of physical process.

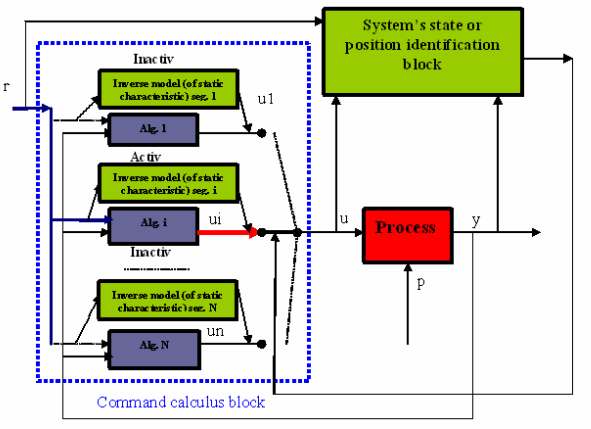

Figure 10. RST control algorithm structure.

IV. ADVANTAGES AND DISADVANTAGES OF THE PROPOSED STRUCTURE

A. Advantages

The main advantage consists in using a simplified and performance operating control structure. Designing procedure is based on classic pole placement and determination of inverse command blocks. Well known procedures are used for dynamic and static models' identification.

Because the global command contains a "constant" component generated by an inverse model command block, the system has good stability margin.

The inverse model command generator can be replaced by a fuzzy logic bloc or neural network that can "contain" human experience about some nonlinear processes.

Due to the fact that the control laws are not very complex, real time software and hardware implementation doesn't need important resources.

B. Disadvantages

The main limitation is that this procedure can be applied only for the processes that permit the construction of the static characteristic.

This structure is very difficult to use for systems with non bijective characteristic and for systems with different functioning regimes.

Another limitation is that this structure can be used only for stable processes. In situations where the process is "running", the direct (feed forward) command is very possible to not have enough flexibility to control it.

The increased number of experiments for determination of the mean static characteristic can be another disadvantage of the structure.

C. Possible Developing

For special situations, the direct command generators (feed forward) included in multi model structure can be constructed as a single general block. This block

compensates the process nonlinearity and allows using simplified control laws in multiple controller structure.

These systems can be easily implemented on PLC structures particularly, and real time control systems, generally.

V. EXPERIMENTAL RESULTS

We evaluate the achieved performances of the proposed control structure using an experimental installation presented in Fig. 11, where the position of an object contained in the vertical tube must be controlled using an air flow generator.

The nonlinear process static characteristic is presented in Fig. 12.

For this, there are selected 12 zones (Fig. 12) for a classic multi model structure (Fig. 1). The models and corresponding area (output %) are: M1: 0-35%, M2: 35-50%, M3: 50-54%, M4: 54-60%, M5: 60-69%, M6: 69-72%, M7: 72-75%, M8: 75-78%, M9: 78-84%, M10: 84-86%, M11: 86-95%, M12: 95-100%. All 12 are first order models. For example, for M1 using Te=0.2 s sampling time and Least Square identification method from Adaptech/WinPIM the model is:

$$M_1 = \frac{0.487180}{1 - 0.79091 q^{-1}}$$

The corresponding controller (for Tracking performances: second order dynamic system with w0=2.0, x=0.95, Disturbance rejection performances: second order dynamic system with w0=1.1, x=0.8, using WinReg) is:

Figure 11. Experimental installation.

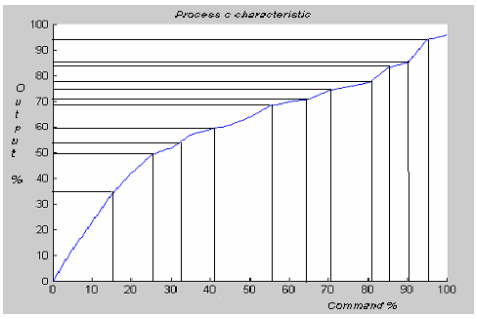

Figure 12. Nonlinear process characteristic.

$$R(q^{-1}) = 0.263281 - 0.179872\, q^{-1}$$
$$S(q^{-1}) = 1.000000 - 1.000000\, q^{-1}$$
$$T(q^{-1}) = 2.052629 - 3.412794 q^{-1} + 1.443573\, q^{-2}$$

The RST control algorithm can be written as follows:

$$u(k) = \frac{1}{s_0}[-\sum_{i=1}^{n_S} s_i u(k-i) - \sum_{i=0}^{n_R} r_i y(k-i) + \sum_{i=0}^{n_T} t_i y^*(k-i)] \quad (9)$$

where *R, S, T* polynomials are presented in relation (3) and n_S, n_R, n_T express the corresponding degrees and also the memory dimension for the software implementation of the algorithm. For example, if $n_R=2$, then three memory locations must be reserved for the process's output: *y(k), y(k-1), y(k-2)*. Respectively, the same rule applies for *u(k)* and $y^*(k)$.

To calculate the corresponding command the controller presented before, there are used 7 multiplication and 7 addition or subtraction operations.

Because the multi models control structure must assure no bump commutations, all of the 12 control algorithms must work in parallel [6]. This condition gives the total number of operations: 12 x 7 = 84 multiplications and 12 x 7 = 84 additions/ subtractions.

For the proposed control structure, with nonlinear blocks there are selected 3 zones Z1: 0-50%, Z2: 50-80% and Z3: 80-100%, presented in Fig. 13.

The models are:

$$M_1 = \frac{0.0964 - 0.19647 q^{-1}}{1 - 1.06891 q^{-1} + 0.22991 q^{-2}}$$
$$M_2 = \frac{0.01297 + 0.05397 q^{-1} + 0.03674 q^{-2}}{1 - 0.76251 q^{-1}}$$
$$M_3 = \frac{0.02187 + 0.05668 q^{-1} + 0.06048 q^{-2}}{1 - 0.93161 q^{-1} + 0.02741 q^{-2} + 0.09863 q^{-3}}$$

In this case, we have computed three corresponding RST algorithms using a pole placement procedure from Adaptech/WinREG platform.

The same nominal performances are imposed to all systems, through a second order system, defined by the dynamics $\omega_0 = 1.25$, $\xi = 1.2$ (tracking performances) and $\omega_0 = 2$, $\xi = 0.8$ (disturbance rejection performances) respectively, keeping the same sampling period as for identification.

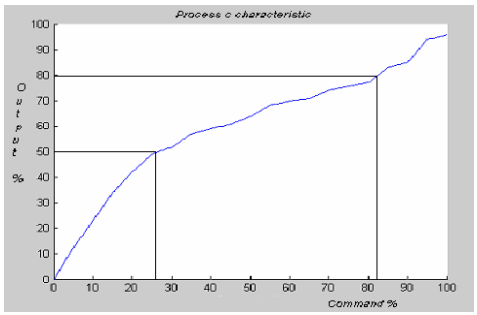

Figure 13. Selection of the three zones of nonlinear characteristic.

$$R_1(q^{-1}) = 1.863259 - 2.027113q^{-1} + 0.520743q^{-2}$$
$$S_1(q^{-1}) = 1.000000 - 0.554998q^{-1} + 0.445002q^{-2}$$
$$T_1(q^{-1}) = 3.414484 - 4.931505q^{-1} + 1.873910q^{-2}$$

$$R_2(q^{-1}) = 2.309206 - 1.624937q^{-1}$$
$$S_2(q^{-1}) = 1.0 - 0.815278q^{-1} - 0.106427q^{-2} - 0.078295q^{-3}$$
$$T_2(q^{-1}) = 9.645062 - 14.928993q^{-1} + 5.968200\,q^{-2}$$

$$R_3(q^{-1}) = 1.72482 - 1.611292q^{-1} - 0.03784q^{-2} + 0.292903q^{-3}$$
$$S_3(q^{-1}) = 1.0 - 0.725187q^{-1} - 0.095205q^{-2} - 0.179608q^{-3}$$
$$T_3(q^{-1}) = 7.192692 - 11.645508q^{-1} + 4.821405\,q^{-2}$$

To calculate the corresponding command for the *C1* controller there are used: 9 multiplications and 9 additions or subtractions, for *C2*: 9 multiplications and 9 additions or subtractions and for C3 11 multiplications and 11 additions or subtractions, giving a total number of 29 multiplications and 29 additions or subtractions.

For the proposed control structure, in addition to the command calculus operation, here is the calculus for the direct command. This depends on the software implementation. For PLC, particular and real time process computers, where (C) code programming can be used, the implementation is:

```
// segment determination
segment = (int)(floor(rdk/10));
// segment gain and difference determination
panta = (tab_cp[segment+1] - tab_cp[segment]) * 0.1;
// linear value calculus
val_com_tr = uk + 1.00 * (panta * (rdk - segment*10.0)
+ tab_cp[segment]);
```

One needs 10 multiplications and 4 additions or subtractions. The total operations number for the proposed structure is: 59 multiplications and 41 additions or subtractions.

It is obvious that the proposed structure needs a diminished number of multiplications if compared to the classic multi model solutions and a comparative value for the number of additions and subtractions. This means that the system with nonlinear compensators is faster or needs simplified hardware and software architecture.

VI. CONCLUSIONS

In this paper there is proposed a multi model control structure which contains, for each model/controller, a nonlinearity compensator.

This solution allows a reduced number of models and a reduced complexity for global structure. The analysis on the advantages and disadvantages of proposed structure is made.

The experimental results, done on laboratory installation, present a case where the proposed structure is a faster solution then the classic multi model structure.

This structure can be easily implemented on PLC and real time process computer.

ACKNOWLEDGMENT

This work was cofinanced from the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013, project number POSDRU/89/1.5/S/56287 „Postdoctoral programs at the forefront of the excellence research in the technologies of the informational society and innovative product and process development", partner University of Oradea and „Automatics, Process Control and Computers" Research Center from University „Politehnica" of Bucharest. (U.P.B. – A.C.P.C.) and Automatic Control and Computers Faculty projects.

REFERENCES

[1] K. S. Narendra and J. Balakrishnan, "Adaptive Control using multiple models", *IEEE Transactions on Automatic Control*, vol. 42, no. 2, Feb., 1997, pp. 171–187

[2] I. D. Landau., R. Lozano and M. M'Saad, *Adaptive Control*, Springer Verlag, London, ISBN 3-540-76187-X, 1997.

[3] L. Ljung, T. Soderstroom, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, Massashusetts, 1983

[4] D. Stefanoiu, J. Culita and P. Stoica, - *Fundamentele Modelarii si Identificarii Sistemelor*, Editura Printech, 2005, ISBN : 973-718-368-1.

[5] G. Tao and P. Kokotovic, *Adaptive control of systems with actuator and sensor nonlinearities*, Wiley, N.Y. 1996

[6] Lupu, D. Popescu, B. Ciubotaru, C. Petrescu, G. Florea ,- *Switching Solution for Multiple Models Control Systems*, MED06, paper WLA2-2, 28-30 June, Ancona, Italy. 2006.

[7] I. Dumitrache, *Ingineria Reglarii Automate*, Politehnica Press, Bucuresti, 2005, ISBN 973-8449-72-3.

[8] J.M. Flaus, *La Regulation Industrielle*, Editure Herms, Paris, 1994, ISBN 2-86601-441-3.

[9] G. Pajunen, "Adaptive control of wiener type nonlinear system", *Automatica*, no. 28, 1992, pp. 781-785

[10] J. Richalet, *Practique de la Commande Predictiv*, Editura Herms, Paris, 1993, ISBN 978-2212115536.

Simulation of Temperature Control in Fermentation Bioreactor for Ethanol Production

MARGINEAN Ana-Maria, MARGINEAN Calin, TRIFA Viorel

Technical University of Cluj Napoca, Romania,
Department of Electrical Machines and Drives, Faculty of Electrical Engineering,
str. Memorandumului nr.28, 400114 Cluj Napoca, Romania, E-Mail:
anamaria.marginean@gmail.com, ignatc@edr.utcluj.ro, trifa@edr.utcluj.ro

***Abstract* – Present paper deals with aspects regarding the simulation of fermentation bioreactor process and fermentation bioreactor control for ethanol production. The bioreactor model was implemented in Matlab Simulink and the results of simulation using different control strategies are presented comparatively. Three types of control strategy are used respectively, PID, Neural Network Model Predictive Controller (NN-MPC) and Nonlinear Auto Regressive Moving Average(NARMA-L2) control strategy.**

Keywords: *fermentation bioreactor, temperature control , PID, Neural Network Model Predictive controller, NARMA-L2 control strategy.*

I. INTRODUCTION

Today's concerning about global warming and the rapid depletion of coal, gas and crude oil reserves enforced the study of alternative fuels as bioethanol.

Bioethanol can be blended at low concentrations with gasoline (usually 10% ethanol and 90% gasoline) or diesel for use in today's vehicles without engines modifications and without affecting vehicle warranty, and is considered to be a sustainable transportation fuel. Alternatively, if bioethanol is used in higher, or 100 % concentrations, but in this case adopted vehicles engines are typically needed.

Starting with biomass harvesting, there are a number of steps to follow until the final product, the ethanol is obtain[1]:

- biomass harvesting – sugar cane, corn, forest residues, municipal waste etc.;
- biomass handling – size reduction step to make biomass easy to handle;
- biomass pretreatment - hemicellulose fraction of the biomass is broken down into simple sugars, and a small part of cellulose is converted to glucose;
- cellulose hydrolysis - the remaining cellulose is hydrolyzed to glucose;
- glucose fermentation - through fermentation the glucose is converted to ethanol;
- ethanol recovery – ethanol is separated from the other components.

Of these processes, present paper is studying the alcoholic fermentation of glucose to ethanol. The bioreactor in which the glucose fermentation takes place is a continuous stirred-tank reactor with constant substrate feed flow and is shown schematically in fig. 1.

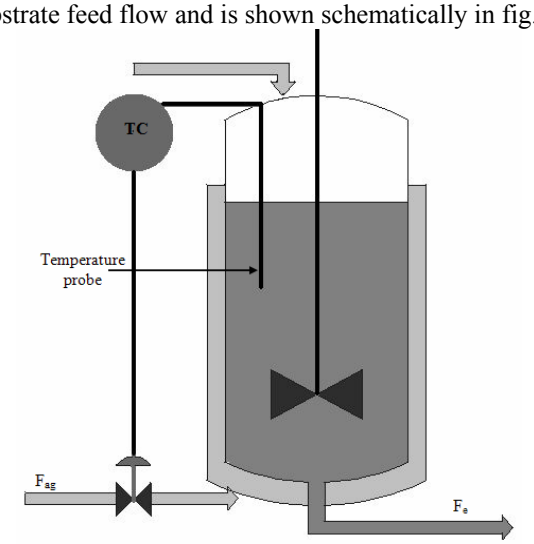

Figure 1. Schematic of fermentation bioreactor.

The three main components of the bioreactor are[2]:

- the biomass as a suspension of yeast fed into the system and evacuated continuously;
- the substrate which is solution of glucose needed in order to feed the micro-organism;
- ethanol as final product evacuated together with other components.

Inorganic salts, which are necessary compounds for the formation of coenzymes, are added together with the yeast.

II. BIOREACTOR MODEL

The kinetic equation used in the bioreactor model represents the Monod modified equations based on Michaelis-Menten kinetics, proposed by Aiba et al., and described by Z.K. Nagy[2].

The mass balances for the biomass is expressed by equation (1) as:

$$\frac{dc_X}{dt} = \mu_X c_X \frac{c_S}{K_S + c_S} e^{-K_P c_P} - \frac{F_e}{V} c_X \quad (1)$$

where c_X represent the biomass(yeast) concentration(h^{-1}) , μ_x is the maximum specific growth rate(h^{-1}), c_S is the glucose concentration(g/l), K_S is constant in the substrate term for growth(g/l), K_p represent constant of growth inhibition by ethanol(g/l), c_P is the product concentration (g/l), F_e is the outlet flow from the reactor(lh^{-1}), V is the volume of the mass of reaction(l).

The mass balance for product is obtain as[2]:

$$\frac{dc_P}{dt} = \mu_P c_X \frac{c_S}{K_{S1} + c_S} e^{-K_{P1} c_P} - \frac{F_e}{V} c_P \quad (2)$$

where μ_P represent the maximum specific fermentation rate(h^{-1}), K_{S1} is the constant in the substrate term for ethanol production(g/l) and K_{P1} is the constant of fermentation inhibition by ethanol(g/l).

Equation (3) represent the mass balance for the substrate:

$$\frac{dc_S}{dt} = -\frac{1}{R_{SX}} \frac{dc_X}{dt} - \frac{1}{R_{SP}} \frac{dc_P}{dt} + \frac{F_i}{V} c_{S,in} - \frac{F_e}{V} c_S \quad (3)$$

where R_{SX} is ratio of cell produced per glucose consumed for growth, R_{SP} is the ratio of ethanol produced per glucose consumed for fermentation, and $c_{S,in}$ represent the glucose concentration in the feed flow.

For the reactor and jacket, the energy balance is described by equations (4) and (5)[2]. In this equations, T_{in} is the temperature of the substrate flow into the reactor(^0C), T_r represent the temperature in the reactor(^0C), r_{O2} is the rate of oxygen consumption($mgl^{-1}h^{-1}$), ΔH_r is the reaction heat generated by the fermentation process, ρ_r and ρ_{ag} is the density of the mass of reaction respectively the density of cooling agent(g/l).

$$\frac{dT_r}{dt} = \frac{F_i}{V}(T_{in} + 273) - \frac{F_e}{V}(T_r + 273) + \frac{r_{O_2} \Delta H_r}{32 \rho_r C_{heat,r}} + \frac{K}{}$$

$$\frac{dT_{ag}}{dt} = \frac{F_{ag}}{V_j}(T_{in,ag} - T_{ag}) + \frac{K_T A_T (T_r - T_{ag})}{V_j \rho_{ag} C_{heat,ag}} \quad (5)$$

Also in the equations (4) and (5), K_T represent the heat transfer coefficient($Jh^{-1}m^{-2}K^{-1}$), A_T is the heat transfer area(m^2) and $C_{heat,ag}$ respectively $C_{heat,r}$ are the heat capacity of cooling agent and of mass of reaction($Jg^{-1}K^{-1}$).

The rate of oxygen consumption is[2]:

$$r_{O_2} = \mu_{O_2} \frac{1}{Y_0} c_X \frac{c_{O_2}}{K_{O_2} + c_{O_2}} \quad (6)$$

with μ_{O2} maximum specific oxygen consumption rate(h^{-1}), K_{O2} constant of oxygen consumption(g/l) and c_{O2} oxygen concentration in the liquid phase(mg/l).

The balance for total volume for reaction medium is:

$$\frac{dV}{dt} = F_i - F_e \quad (7)$$

Equation (1) to (7) were implemented in Matlab Simulink in order to study the dynamic behavior of the system and testing different control strategies.

III. BIOREACTOR MODEL IMPLEMENTATION

The implementation of bioreactor model using Matlab Simulink is presented in Fig. 2.

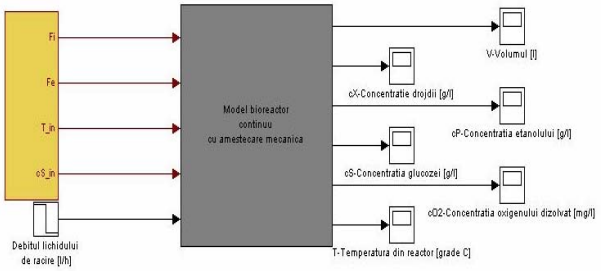

Figure 2. Implementation of the bioreactor model.

Each of the equations describing the model was implemented using Function Blocks from Simulink.

Because the temperature in the reactor depends of the flow of cooling agent, the first step in our analysis was to apply a step change in the flow of cooling agent in order to study the influence on the reactor temperature. The applied step change and the evolution of the temperature in the reactor are depicted in Fig. 3.

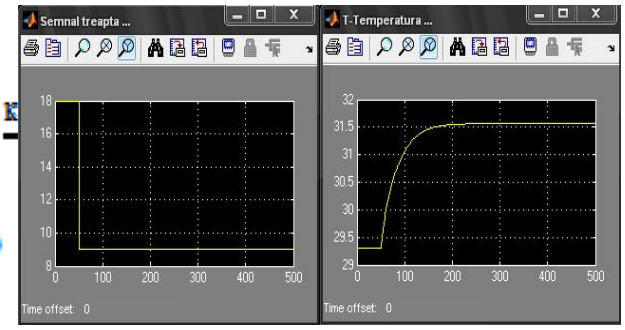

Figure 3. Step change and the reactor temperature evolution.

In order to study the dynamic behavior of the bioreactor system a step change with 2^0C in inlet flow temperature was applied, which can occur due to the ambient temperature[2]. The results are presented in figure 4.

Analyzing the results of the simulation, one can observe that the effect of the change in the inlet

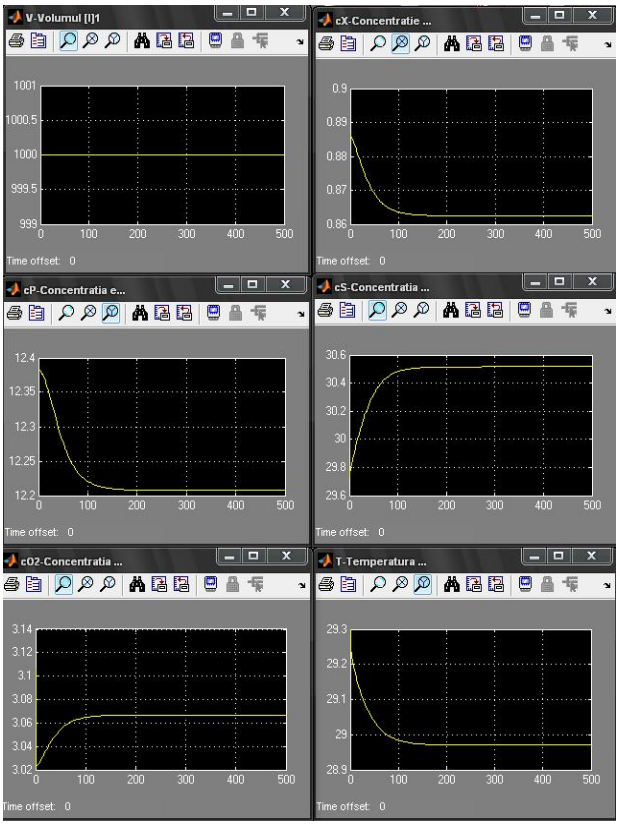

Figure 4. Dynamic response of the bioreactor in the case of step change in the temperature input flow.

temperature have a major impact on the ethanol concentration and can be considered as a major disturbance in the system.

III. TEMPERATURE CONTROL STRATEGIES

Three control strategies were taken into account, namely proportional-integral-derivative(PID) control, Neural Network Model Predictive Controller (NN-MPC) and Nonlinear Auto Regressive Moving Average (NARMA-L2) control strategy.

A. Neural Network Model Predictive Controller

The neural network predictive controller uses a neural network model of a nonlinear plant in order to predict future plant responses. The neural network plant model is trained offline using any of the training algorithms such as Levenberg-Marquard, Bayesian regularization, BFGS Quasi-Newton, conjugate gradient and gradient descent [6].

The controller, however, requires a significant amount of online computation, because an optimization algorithm is performed at each sample time to compute the optimal control input.

The first step in model predictive control is to determine the neural network plant model (system identification). In this step, the neural network is train to represent the forward dynamics of the plant. The prediction error between the plant output and the neural network output is used as the neural network training signal[5].

In the second step, the plant model is used by the controller to predict future performance. The entire process is depicted in Fig. 5.

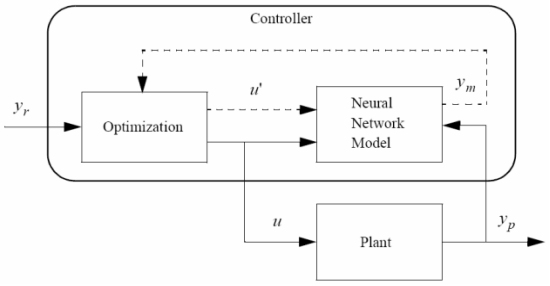

Figure 5. Diagram of Model Predictive Control Process.

The controller consists of the neural network plant model and the optimization block. The optimization block determines the values of u' that minimize J, and then the optimal u is input to the plant[5].

B. Neural Network Feedback Linearization Control(NARMA-L2)

By canceling the nonlinearities, this type of control transform nonlinear system dynamics into linear dynamics. In Fig. 6 a block diagram of the NARMA-L2 controller is presented, with blocks labeled TDI representing tapped delay lines that store previous values of the input[4].

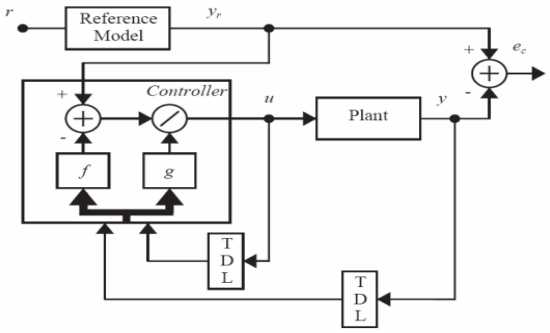

Figure 6. Block diagram of the NARMA-L2 controller.

This type of controller requires the least computation, and represent a rearrangement of the neural network plant model, which is trained offline, in batch form.

IV. SIMULATION RESULTS

Using Neural Network Toolbox™ from Matlab, the NN-MPC and NARMA-L2 controllers were implemented. The Simulink model for neural network control of fermentation bioreactor is presented in Fig. 7, in case of NARMA-L2 controller, with the controller as an independent block.

The network architecture and the NN-MPC respectively the NARMA-L2 controllers parameters are summarized in Table 1.

For both controllers, the network used in these simulations have 4 neurons in the input layer, 1 neuron in the output layer, two hidden layers with 8 neurons of hyperbolic tangent activation function and 1 neurons with linear activation function.

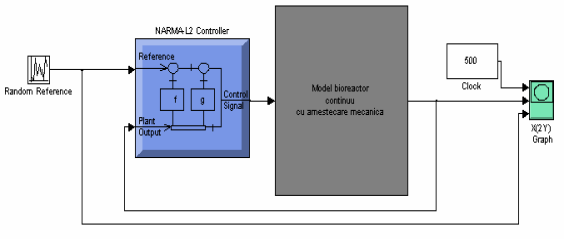

Figure 7. Simulink model for neural network control of fermentation bioreactor in case of NARMA-L2 controller.

Parameters for controllers as cost horizon, control horizon, control weighting factor, search parameters, iterations per sample time and the size of hidden layer are determined using trial and error method.

TABLE 1. Type sizes for camera-ready papers.

Parameters	*NN-MPC*	*NARMA-L2*
Train epochs	200	200
Train function	Bayesian Regulation	Bayesian Regulation
Train data	Normalized	Normalized
Cost horizon	5	-
Control horizon	2	-
Minimization routine	csrchbac	
Control weighting factor	0.05	-
Search parameter	0.001	-

Regarding the PID controller, a two-degree-of-freedom(2DOF) PID controller in parallel form with automatic tuning options, was used.

The best results of the simulation regarding the response of temperature controllers to set point change are presented in Fig. 8. For comparison the results obtained with a classical PID controller type are also presented in the same figure. As one can see by analyzing the results, settling time of NARMA-L2 controller response is much shorter than PID and NN-MPC controllers response in the face of set point changes.

V. CONCLUSIONS

The mathematical model for bioreactor used in glucose fermentations is presented and implemented using Matlab Simulink. Also Neural Network Model Predictive Controller (NN-MPC) and Nonlinear Auto Regressive Moving Average(NARMA-L2) control strategy are presented and used for temperature control of bioreactor. The results of the study are presented comparatively with the results obtain using a classical PID controller. From this three controllers, the best response to set point change was offered by the NARMA-L2 controller.

ACKNOWLEDGMENTS

This paper was supported by "Project for development of doctoral studies in advanced technology-PRODOC", project co-funded by the European Social Fund through the Sectorial Operational Program Human Resources 2007-2013.

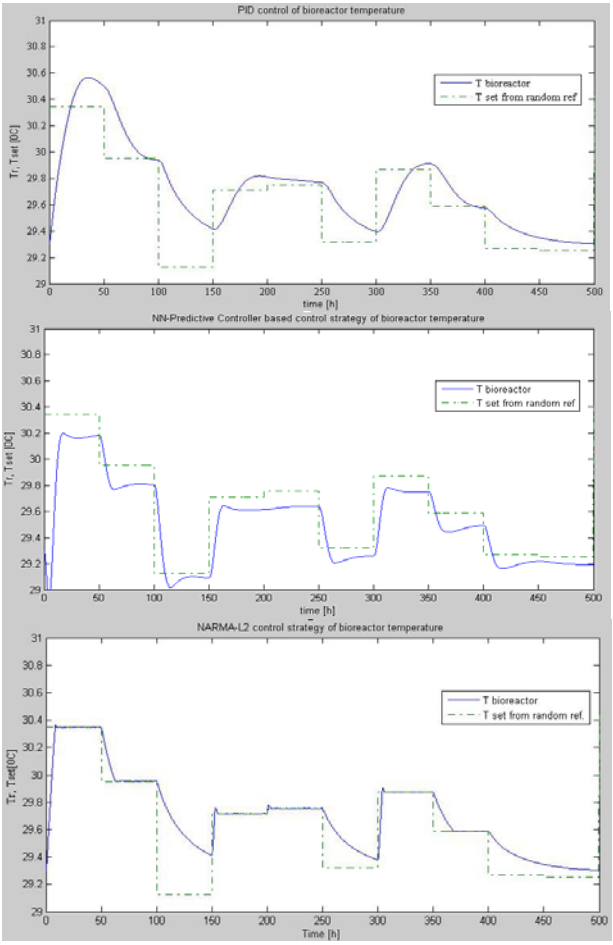

Figure 8. Simulation results with PID, NN-MPC, and NARMA-L2 control of the process.

REFERENCES

[1] http://www1.eere.energy.gov/biomass/m/abcs_biofuels.html.

[2] Z.K. Nagy, "Model based control of a yeast fermentation bioreactor using optimally designed artificial networks" Elsevier, Chemical Engineering Journal 127(2007)95-109.

[3] A. Assadzadeh, S.S. Jamuar, "Development and simulation of biochemical reactor by using Matlab", 12th International Conference on Computer Modelling and Simulation, 2010.

[4] Mete, T., Ozkan, G., Hapoglu, H. and Alpbaz, M. (2010), Control of dissolved oxygen concentration using neural network in a batch bioreactor. Comput. Appl. Eng. Educ.. doi: 10.1002/cae.20430, 2010.

[5] M. Beale, M. Hagan, H. Demuth "Neural Network Toolbox™ User's Guide", The Math Works Inc. 2012.

[6] B. ZareNezhad, A. Aminian, "Application of the neural network-based model predictive controllers in nonlinear industrial systems. Case study.", Journal of the University of Chemical Technology and Metallurgy, 46, 1, 2011, Sofia, Bulgaria, pp. 67-74.

Average Torque Control of an 8/6 Switched Reluctance Machine for Electric Vehicle Traction

PETRUS Vlad[1, 2], POP Adrian-Cornel[1, 2], GYSELINCK Johan[2], MARTIS Claudia[1], IANCU Vasile[1]

[1] Technical University of Cluj-Napoca, Romania,
Department of Electrical machines and drives, Faculty of Electric Engineering,
Memorandumului Street, No. 28, 400114, Cluj-Napoca, Romania,
vlad.petrus@mae.utcluj.ro

[2] Université Libre de Bruxelles, Belgium,
Department of Bio-, Electro- and Mechanical Systems, Faculty of Applied Sciences,
Avenue Franklin Roosevelt, 50, B-1050, Brussels, Belgium,
johan.gyselinck@ulb.ac.be

Abstract ***– This paper presents two average torque control techniques implemented on a 30kW peak-power 8/6 switched reluctance machine (SRM) used for electric vehicle traction: a feed-forward open-loop technique and a closed-loop technique. The computation of the control drive parameters used by the feed-forward open-loop control scheme is briefly described together with simulation and experimental results. A method of on-line average torque estimation, making possible the implementation of the closed-loop average torque control method is described. Measurements on the test bench are presented and a discussion is made regarding the advantages of the closed-loop technique.***

Keywords: ***Switched Reluctance Machine; Average Torque Control; Closed-loop; Vehicle Traction;***

I. INTRODUCTION

The undeniable advantages of the SRM, in terms of simplicity, robustness, fault tolerance, low maintenance and price, make it a candidate with real chances on the market of electric traction. The salient structure of the machine and its highly non-linear behavior lead to important torque ripple, representing the main drawback of the machine.

At low speeds, the low-frequency torque ripple generates oscillations that may excite resonant frequencies of the drive train, making the vehicle undrivable. At higher speeds the frequency of the ripple increases, causing high-frequency acoustic noise, disturbing for the human ear.

To overcome these drawbacks, different torque-ripple minimization techniques were proposed in the literature regarding the design and control of SRMs.

In Fig. 1 is illustrated a basic feed-forward torque control based on on-line or off-line computed tables of the phase current, turn-on and turn-off angles as functions of electromagnetic torque and speed. In traction drives, where the variation of the DC-bus voltage plays an important role in the stability of the controller, the before mentioned tables have an extra dimension represented by this DC-bus voltage. The extra dimension increases the computational effort and requires more expensive hardware. The influence of the firing angles on the performance of the average torque controller was studied in several papers [1-5].

An expression to calculate the turn-on angle was proposed by Bose [6]. He considered that the current has to reach its reference value in the point where the rotor and stator poles start to overlap and the torque producing region begins. The turn-off angle can be calculated with the expression proposed by Gribble [7].

In [8] and [9] an on-line efficiency optimization scheme is presented. Firing angles are computed on-line using the relations of Bose and Gribble. In steady state operation the initial selection of the angles is fine-tuned by means of an algorithm that minimizes the input power of the drive. Due to dynamically changing operating point, this method is not suited for traction drives.

In [1, 2, 10, 11], a method of improving both the torque ripple and the efficiency is proposed with an on-line calculation of the firing angles. The turn-on angle can be calculated with the expression proposed by Bose. Empirically was determined that the optimal turn-off angle for the highest efficiency has to be chosen so that the fluxes of two neighboring phases are equal to half of the peak flux-linkage on their intersection angle.

The multitude of simplifying hypothesis makes these methods inaccurate. So, the need of optimization techniques taking into account all non-linearities of the SRM aroused. As a starting point the maximization of the average torque per ampere was aimed [6, 12], but soon the need of a secondary objective like efficiency or torque-ripple minimization has been acknowledged [7, 13]. Optimization of firing angles with multiple secondary objectives using weight factors are found in [14-16].

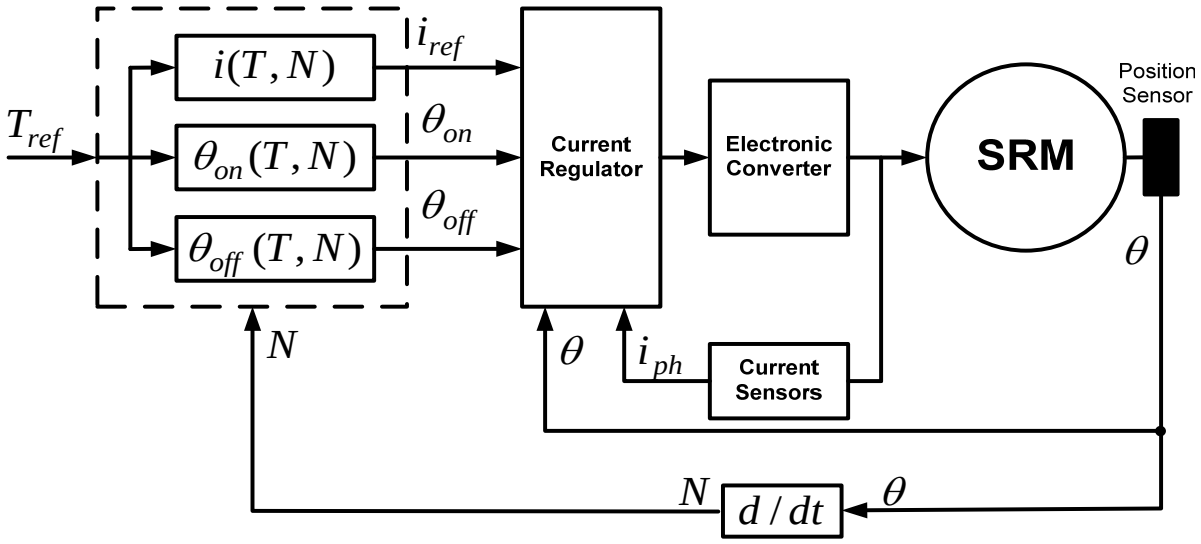

Figure 1. Diagram of feed-forward open-loop average torque control.

This paper presents two average torque control techniques implemented on 30kW peak-power 8/6 switched reluctance machine used for electric vehicle traction: a feed-forward open-loop technique and a closed-loop technique. In section II, the computation of the control drive parameters used by the feed-forward open-loop control scheme is briefly described together with simulation and experimental results. Next section describes a method of on-line average torque estimation, making possible the implementation of the closed-loop average torque control method. Section IV presents measurements on the test bench and the advantages of the closed-loop technique. Finally conclusions are drawn and future directions for the work are suggested.

II. DRIVE PARAMETER COMPUTATION

An optimization routine using a multi-objective minimization function is defined in [5, 16]. The torque per ampere, the torque ripple and the efficiency were considered as performance indicators used to define a function were weight factors for each performance parameter are considered. The output of the optimization routine is represented by the reference current, the turn-on and turn-off angles. The minimization routine was conducted considering a weight factor of 75% for the torque ripple and a weight factor of 25% for the efficiency. The variation with torque and speed of the output quantities is presented in Fig. 2, Fig. 3 and Fig. 4.

The data coming out from the optimization routine in its raw form presents local minima or maxima. The non-monotonic variation of the quantities can lead to malfunctions in the controller and/or unwanted noise. Thus the quantities have to be passed through a smoothening procedure.

The reference current increases with torque and slightly with speed. The turn-on and turn-off angles have to be advanced with the increase of speed and torque. The turn-on angle has to be chosen in such a manner that the current reaches its reference value in the point where the slope of the inductance starts to increase and torque production is optimum and the turn-off angle has to be chosen such that the current becomes zero before the aligned position to avoid negative torque production.

The waveforms of the phase currents, the phase fluxes and the electromagnetic torque at a reference value of torque of 90Nm and a base speed of 1500rpm presented in Fig. 5. The reference current is 104.37A, and the turn-on angle and the turn-off angle are 35.39° and 53.57°. The maximum and the minimum registered torque values are 83.7Nm and 103Nm. The absolute torque ripple is 19.3Nm, while the relative ripple representing 21% of the average torque.

In Fig. 6 the waveform of the phase currents, the voltage on one phase and the electromagnetic torque are illustrated. An offset between the imposed value of the torque set at 30Nm and the average measured value is noticed. These often spotted deviations appear due to the variation of the machine parameters with the environmental conditions, the variations of the DC-bus voltage and due to the off-line calculated control variables in ideal conditions.

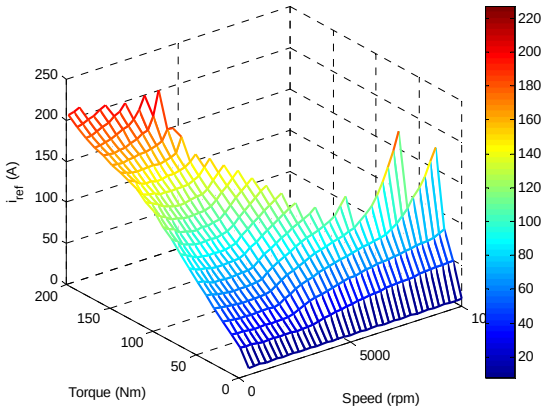

Figure 2. Reference current vs. torque and speed.

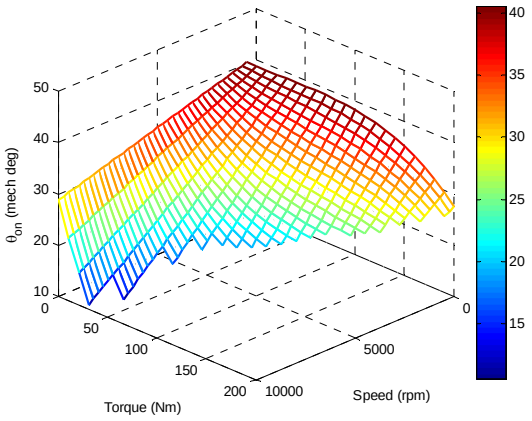

Figure 3. Turn-on angle vs. torque and speed.

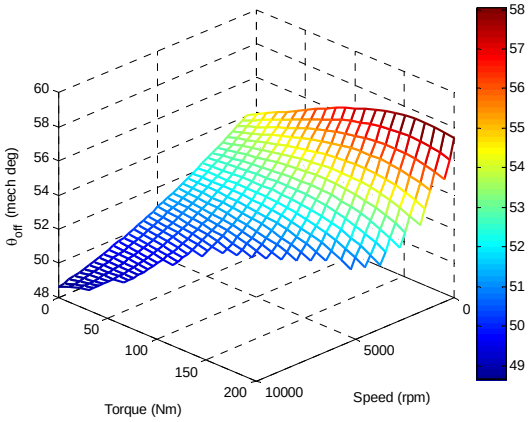

Figure 4. Turn-off angle vs. torque and speed.

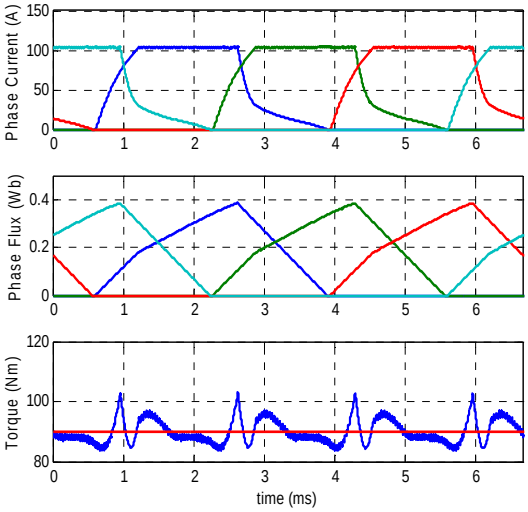

Figure 5. Simulated waveforms of phase currents, phase fluxes and instantaneous torque for a reference torque of 90Nm at 1500rpm.

To avoid these unwanted deviations between the torque reference and the torque output, the need of adjusting the variables in a closed-loop control was emphasized [17].

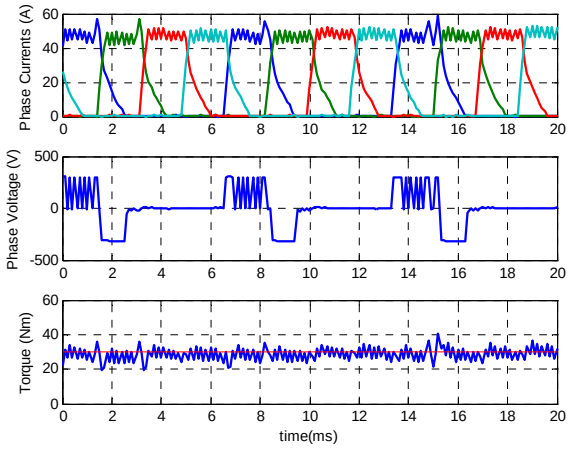

Figure 6. Measured waveforms of phase currents, voltage on one phase and instantaneous torque for a reference torque of 30Nm at 1500rpm.

III. ON-LINE AVERAGE TORQUE ESTIMATION

The closed-loop average torque control technique requires on-line average torque estimation. The on-line average torque estimation of an SRM, as the instantaneous-torque estimation is not a facile task [17-19]. For SRMs with the number of phases higher than three, two or more phases can participate at once to the average torque production.

The average electromagnetic torque estimation in switched reluctance machine requires the knowledge of the phase flux. The phase flux can be estimated from the terminal quantities, phase current, i_{ph}, and phase voltage, u_{ph}, and the phase voltage equation:

$$\frac{d\lambda_{ph}}{dt} = u_{ph} - R_{ph} i_{ph} \tag{1}$$

with R_{ph} the phase resistance. The quantity of energy converted from electrical energy, W_{elec}, to mechanical energy , W_{mech}, corresponding to the area of the co-energy loop, can be further calculated:

$$W_{mech} = \oint i_{ph} \frac{d\lambda_{ph}}{dt} dt \tag{2}$$

The average electromagnetic torque can be calculated using the torque equation:

$$T_{avr} = \frac{N_{ph} N_r}{2\pi} W_{mech} \tag{3}$$

If the torque of only one phase is estimated, the on-line computational effort and the need of extra and expensive hardware will be reduced. Using the terminal quantities of only one phase will assume a constant torque over an entire electric cycle, meaning four strokes in the given case. This limits the speed and the quality of the response of the controller.

The structure of the average torque estimator, Fig. 7, shows that the calculation of torque is based on the energy balance. The integrator has to be reset each time the current becomes zero to avoid drift on the electrical

energy. A sample and hold block is used to determine the mechanical energy.

Fig. 9 and Fig. 10 show the reaction of the estimator to a step change and to a ramp change in the reference torque. In both figures, from top to bottom, are the phase current, the electric energy and the instantaneous torque and the average estimated torque. The estimated average torque shows a delay of one electrical period. If the estimation is made on every phase, the delay can be reduced to one fourth of the electrical period, but this will increase the computational effort.

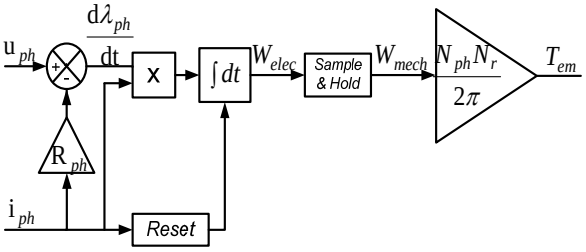

Figure 7. Structure of the average torque estimator.

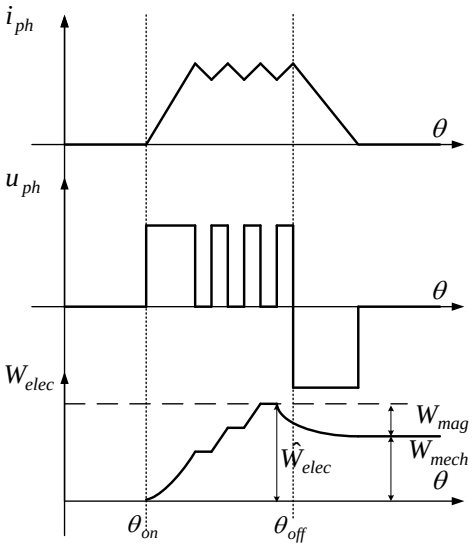

Fig. 8. Typical quantities for the on-line torque estimator.

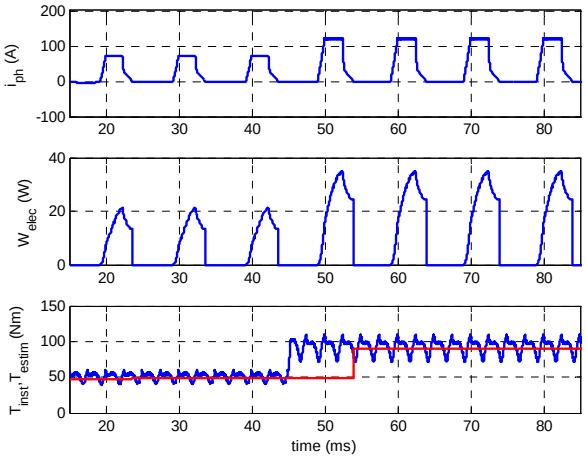

Figure 9. Reaction of the average torque estimator at a step change in the average torque.

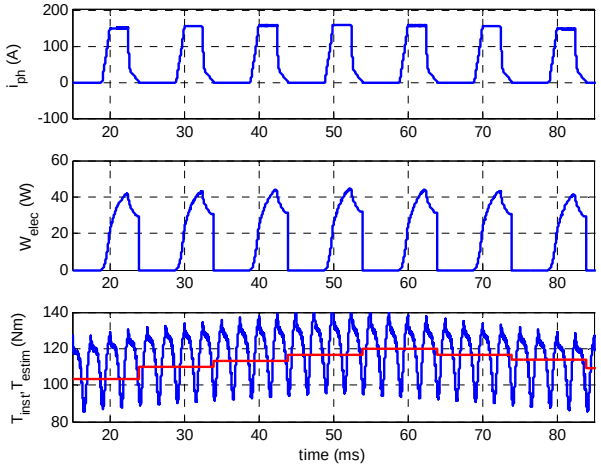

Figure 10. Reaction of the average torque estimator at a ramp change in the average torque.

III. CLOSED-LOOP AVERAGE TORQUE CONTROL

The open-loop controller is sensitive to variations of machine parameters, which are influenced by environmental conditions. The control variables are estimated on-, but usually off-line in ideal conditions. Thus, the experimental drive output is susceptible deviate from the reference torque value. To avoid these unwanted deviations between the torque reference and the torque output, average torque is tracked closed-loop control, adjusting the control variables.

The torque error treated trough a proportional-integral regulator is added to the reference torque for the control variables becomes the sum of the two, Fig 11.

The experimental set-up developed for the validation of the theoretical approach is presented in Fig. 12. The SRM loaded with a DC machine is fed via an asymmetric H-bridge based converter.

A dSPACE 1103 prototyping platform is used to implement the control logic. A 1024-pulse incremental encoder was employed for rotor position measurement. The tests are conducted imposing the maximum switching frequency at 10kHz.

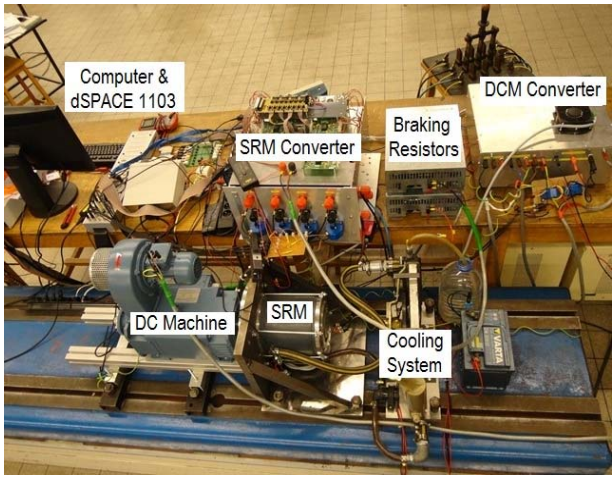

Figure 12. Experimental set-up.

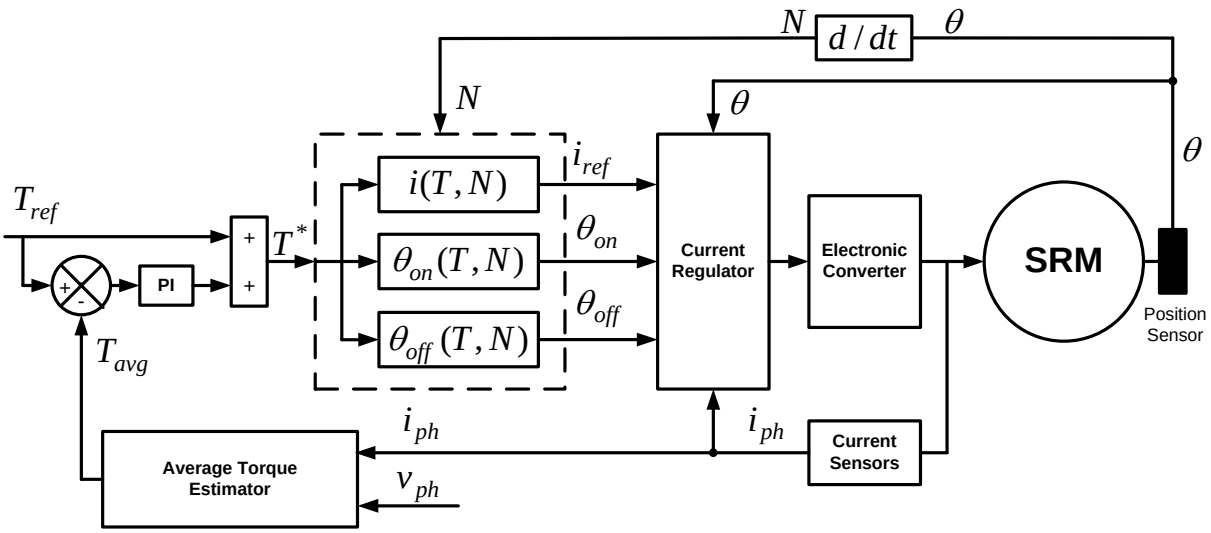

Figure 11. Diagram of closed-loop average torque control.

In Fig. 13, the measured waveforms of the phase currents together with the instantaneous torque are illustrated. The speed was set at 1500rpm and the reference value of torque was 30Nm. It can be noticed that the deviation of the mean value of torque from its reference value was eliminated. Changes in the operation conditions as variations in the DC-bus voltage, variations of the resistance with the temperature, manufacturing imperfections and inaccuracies in the off-line ideal estimation of the control parameters are compensated. The unwanted oscillations caused by the transition from instantaneous torque control techniques used at low speed to average toque control technique used at higher speeds can be eliminated by the closed-loop torque control method [18].

In Fig.14 the response of the controller at different changes in the torque reference it is presented. In the first situation the reference is maintained constant at 20Nm. In the second, the reference varies linearly from 0 to 20Nm in 2 seconds and in the third case, the reference is step changed from 0 to 20Nm. The presented results demonstrate the robustness of the controller under steady-state and dynamic conditions.

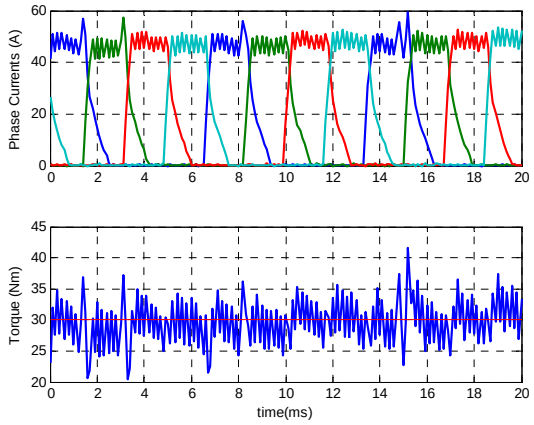

Figure 13. Measured waveforms of phase currents and instantaneous torque for a reference torque of 30Nm at 1500rpm with closed–loop control.

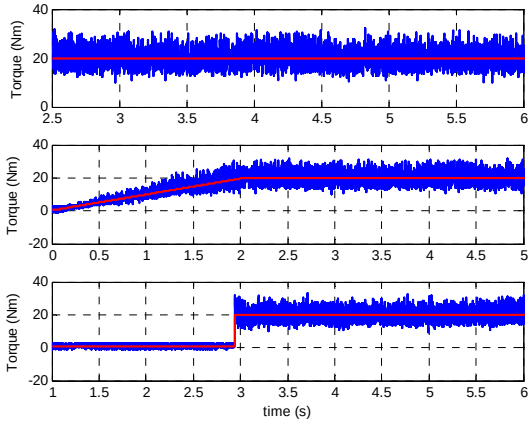

Figure 14. Measured waveforms instantaneous torque for a constant, a ramp varying and step varying reference at 1500rpm with closed–loop control.

III. CONCLUSIONS

In this paper, two average torque control techniques were implemented on 30kW peak-power 8/6 switched reluctance machine used for electric vehicle traction: a feed-forward open-loop technique and a closed-loop technique. The computation of the control drive parameters used by the feed-forward open-loop control scheme was briefly described. Simulation and experimental results were presented emphasizing the need of the closed-loop control. A method of on-line average torque estimation was described, making possible the implementation of the closed-loop average torque control method. Measurements on the test bench and the advantages of the closed-loop technique were presented in the end. The precision and the robustness of the controller under steady state and dynamic conditions was illustrated and discussed.

The proven robustness and the undeniable capabilities of the closed-loop average torque controller make it suitable for high dynamic, wide operating range applications such as electric vehicle traction.

ACKNOWLEDGEMENT

This research was financially supported in the frame of the project "Doctoral studies in engineering sciences with purpose to develop knowledge-based society – SIDOC (abbreviation from Romanian)", Contract POSDRU/88/1.5/S/60078.

REFERENCES

[1] C. Mademlis and I. Kioskeridis, "Performance optimization in switched reluctance motordrives with online commutation angle control," IEEE Transactions on Energy Conversion, vol. 3, pp. 448-457, 2003.

[2] I. Kioskeridis and C. Mademlis, "Maximum efficiency in single-pulse controlled switched reluctance motor drives," IEEE Transactions on Energy Conversion, vol. 20, pp. 809 - 817, 2005.

[3] C. S. Dragu and R. Belmans, "Optimal firing angles control for four-quadrant operation of an 8/6 SRM," 10th European Conference on Power Electronics and Applications, 2003.

[4] C. S. Dragu and R. Belmans, "Four-quadrant control of an 8/6 switched reluctance motor," 4th International symposium on advanced electromechanical motion systems-ELECTROMOTION, vol. 2, pp. 455-460, 2001.

[5] A.-C. Pop, V. Petrus, J. Gyselinck, C. S. Martis, and V. Iancu, "On the firing angles control of a 8/6 switched reluctance machine," Journal of Electrical and Electronics Engineering, vol. 4, pp. 189-194, 2011.

[6] B. K. Bose, T. J. E. Miller, P. M. Szczesny, and W. H. Bicknell, \Microcomputer control of switched reluctance motor," IEEE Transactions on Industrial Applications, vol. IA-22, pp. 708-715, 1985.

[7] J. J. Gribble, P. C. Kjaer, and T. J. E. Miller, "Optimal commutation in average torque control of switched reluctance motors," IEE Proceedings - Electric Power Applications, vol. 146, pp. 2-10, 1999.

[8] B. Blanque, J. I. Perat, P. Andrada, and M. Torrent, "Improving efficiency in switched reluctance motor drives with online control of turn-on and turn-o angles," European Conference on Power Electronics and Applications, 2005.

[9] P. Andrada, B. Blanque, J. I. Perat, M. Torrent, E. Martinez, and J. A. Snchez, "Conmparative efficiency of switched reluctance and induction motor drives for slowly varying loads," ICREPQ06, 2006.

[10] C. Mademlis and I. Kioskeridis, "Smooth transition between optimal control modes in switched reluctance motoring and generating operation," International Conference on Power Systems Transients, IPST07, 2007.

[11] C. Mademlis and I. Kioskeridis, "Four-quadrant smooth torque controlled switched reluctance machine drives," IEEE Power Electronics Specialists Conference, PESC 2008, pp. 1216-1222, 2008.

[12] R. Orthmann and H. P. Schoner, "Turn-o angle control of switched reluctance motors for optimum torque output," Fifth European Conference on Power Electronics and Applications, vol. 6, pp. 20-25, 1993.

[13] P. C. Kjaer, P. Nielsen, L. Andersen, and F. Blaabjerg, "A new energy optimizing control strategy for switched reluctance motors," IEEE Transactions on Industry Applications, vol. 31, pp. 1088-1095, 1995.

[14] A. M. Omekanda, "A new technique for multidimensional performance optimization of switched reluctance motors for vehicle propulsion," IEEE Transactions on Industry Applications, vol. 39, pp. 672-676, 2003.

[15] J. H. Fisch, Y. Li, P. C. Kjaer, J. J. Gribble, and T. J. E. Miller, "Pareto-optimal firing angles for switched reluctance motor control," Second International Conference On Genetic Algorithms in Engineering Systems: Innovations and Applications, GALESIA 97, pp. 90-96, 1997.

[16] A.-C. Pop, V. Petrus, J. Gyselinck, C. S. Martis, and V. Iancu, "Wide-speed range control strategy for an 8/6 switched reluctance machine," ACEMP & Electromotion Joint Conference, 2011.

[17] R. B. Inderka", R. W. D. Doncker, "High Dynamic Direct Average Torque Control for Swiched Reluctance Drives," Thirty-Sixth IAS Annual Meeting, 2001, 3, 2111 – 2115

[18] R. B. Inderka", R. W. D. Doncker, "High Dynamic Direct Average Torque Control for Swiched Reluctance Drives," IEEE Transactions on Industry Applications, 2003, 39, 1040 – 1045

[19] I. Boldea, "Variable Speed Generators," CRC Press, 2005

Finite Element Based Multiphysics Optimal Design of Switched Reluctance Motors Used in Electric Vehicles Propulsion

POP Adrian-Cornel[1, 2], PETRUS Vlad[1, 2], GYSELINCK Johan[2], MARTIS Claudia[1], IANCU Vasile[1]

[1] Technical University of Cluj-Napoca, Romania,
Department of Electrical machines and drives, Faculty of Electric Engineering,
Memorandumului Street, No. 28, 400114, Cluj-Napoca, Romania,
adrian.pop@mae.utcluj.ro

[2] Université Libre de Bruxelles, Belgium,
Department of Bio-, Electro- and Mechanical Systems, Faculty of Applied Sciences,
Avenue Franklin Roosevelt, 50, B-1050, Brussels, Belgium,
johan.gyselinck@ulb.ac.be

Abstract *– In this paper the authors present a FE-based, multiphysics design optimisation procedure for a high power density switched reluctance machines used in HEV/EV propulsion. Two different approaches are considered regarding the initial design. In the first case the topology is obtained using a pre-sizing routine while taking into account the requirements of an electric motor used for HEV/EV, whereas in the second case the performance of the multiphysics optimisation routine is assessed starting from an existing SRM and trying to increase its torque density without modifying the outer envelope nor the cooling circuit. Three different types of constraints are embedded in the optimisation routine. The electromagnetic ones are given by the magnetic flux density (both in the stator pole and stator back iron) whereas the heat generation (thermal constraint) is controlled by means of the current density. Furthermore, at each evaluation the circumferential mode frequency for the pulsating vibration mode m=2 is calculated and lower boundaries are considered for limiting the deterioration of the NVH properties*

Keywords: *SRM, FEM, NVH, HEV, optimal design, multiphysics.*

I. INTRODUCTION

The main drawbacks of the Switched Reluctance Machine (SRM) [1] - namely, noise vibration and harshness problems (NVH), low efficiency and high torque ripple can be overcome either at the machine design level, or from the control technique that is used, as it is done in [2] As for the design, extensive work has been conducted for determining the influence of various parameters on the performance indicators.

Regarding the rotor, the studies reported in [3], [4] have shown that for radially-laminated SRMs increasing the rotor pole width has no direct influence on the peak torque. As for the stator dimensions, results are reported considering the following degrees of freedom (DoFs), yoke thickness [5], stator pole width, internal stator diameter, [4], [5], [6] and the stator pole shapes [7]. The overall conclusions are that the torque increases with the yoke and with the other dimensions up to an optimal value. It is well known that the most effective method to increase the torque is by reducing the airgap, down to the lowest limit allowed by the manufacturing technology [4], [6].

Most of the reported work shows the influence of the geometry modification only on the electromagnetic torque, without tracking how other quantities like iron or copper losses change when the developed torque increases, with few exceptions such as [8]. Moreover most of the authors limit their study to the most common three phase, 6/4 SRM, and only brief accounts are made on higher number of phases.

In this paper the authors present a FE-based, multiphysics design optimization procedure for a high power density switched reluctance machine used in HEV/EV propulsion. Two different approaches are considered regarding the initial design. In the first case it is obtained using a pre-sizing routine [10]-[12] and taking into account the requirements of an electric motor used for HEVs/EVs, whereas in the second case the performance of the multiphysics optimization routine is assessed starting from an existing four-phase, 8/6 SRM, and trying to increase its torque density without modifying the outer envelope nor the cooling circuit. Three different types of constraints are embedded in the optimization routine. The electromagnetic ones are given by the magnetic flux density (both in the stator pole and stator back iron) whereas the heat generation (thermal constraint) is controlled by means of the current density. Furthermore, at each evaluation (of the minimization routine) the circumferential mode frequency for the pulsating vibration mode m=2 is analytically calculated and lower boundaries are considered for limiting the deterioration of the NVH properties.

The paper is structured in five sections. In Section II the used SRM topologies are briefly presented, which in Section III are optimized by means of the multiphysics optimal design. The various optimization scenarios that are considered along with comprehensive results obtained with the two topologies are provided in Section IV. Finally, the conclusions are drawn in the last Section along with suggestions for future work.

II. INITIAL DESIGN

Two different approaches are considered for validation of the optimization routine. In the first case, a (new) motor is designed starting from the output power equation expressed in terms of bore diameter, stack length, speed, and magnetic and electric loadings. Using the imposed quantities (phase voltage, base speed) given by the base requirements for an EV [9], as well as some reasonable values for the desired efficiency, energy conversion ratio, magnetic flux density in the airgap, the bore diameter can be determined [10] and further the other main dimensions of the machine are computed, mainly based on geometrical formulas, as well as experience-based empirical relations [11], [12]. As for the electrical circuit parameters the number of turns per phase is calculated in order to fully determine the data required for building the FE model. The output of the pre-sizing routine [11] which is further referred to as SRMPRE-S is used in the multiphysics optimal design. It is a four-phase 8/6 SRM with the geometry and electrical circuit parameters provided in the initial design (ID) column in Table 2.

According to [13] for SRMs with a current density in the winding less than only 16A/mm^2 a circular tube of coolant fluid is needed around the end-coils for ensuring proper operation. However, if longitudinal channels for the coolant agent are considered, the current density can be increased to 25A/mm^2 or even more if a high convective heat transfer coefficient is ensured by the cooling circuit.

The second motor under study is an existing SRM used in hybrid drive-trains [14], which will be further referred to as SRMPPT. It is a four-phase 8/6 SRM capable of developing 200Nm peak torque and delivering around 15kW in continuous operation up to 10000rpm.

Torque density ($T_\rho = V_{Fe}/T_{av}$, where V_{Fe} is the volume of iron and T_{av} is the average torque) is used as objective function, unless otherwise stated. In order to obtain higher torque for less iron volume, $1/T_\rho$ is minimized using a function which employs the Nelder-Mead downhill simplex algorithm [20].

III. MULTIPHYSICS OPTIMAL DESIGN

A. Introduction

Summarizing all the optimization objectives one may conclude that when the outer diameter and airgap width are constraints, then the available DoFs (considering only the geometrical dimensions) are the height of the poles, yokes width and poles width.

Taking the bore diameter as DoF and automatically adjusting the rest of the geometrical dimensions in the radial direction, so as to maintain the outer diameter fixed, the poles height is indirectly modified. Adding the rotor back iron radius as well as pole arcs as DoFs in the optimisation all the above mentioned parameters are exploited, except the stator back iron thickness. As stated in [1] the optimum pole arcs are a compromise between various conflicting requirements. In [15] it was reported that the optimal values lie within the so-called, feasible triangle. Mathematically it is described by (1) and it is considered in the optimization as an additional constraint in the geometry generation routine (Fig. 4).

$$\begin{cases} \beta_r \geq \beta_s \\ \min(\beta_r, \beta_s) \geq \theta_s \\ \beta_s < 2\pi / N_R - \beta_r \end{cases} \quad (1)$$

In the proposed multiphysics procedure for increasing the electromagnetic torque developed by a SRM, both the outer diameter and stack length are kept constant. Other constraints that are embedded in the optimization routine are chosen as follows:

At each iteration the natural mode frequencies are calculated and as constraint is considered the value of the frequency for the second mode shape (m=2). In order to avoid the resonance with the fundamental (which facilitates the noise production), the second mode eigen-frequency has to exceed the fundamental frequency which corresponds to the waveforms of radial force which characterize the SRM turning at maximum speed.

Temperature is kept in the same range by limiting the increase in the heat generation. The copper losses and current density are kept constant (slot area is adjusted in function of the optimisation scenario). Iron losses augmentation is limited as the stator yoke thickness is not reduced more than (20-40) % of the half of the stator pole width.

In Table 1 the design variables with the index 0 represent the value at the evaluation zero (initial design) whereas Δ_{dof} is factor less than 1.

B. Optimisation Algorithm

a. Computation of the design variables

After running several calculations it was found that in order to maximize the torque, the tendency of the minimization routine is to increase the current density, by decreasing the coil area.

TABLE 1. The design variables and their variation range.

Design Variable	Lower/Upper Boundary
Stator pole arc, β_s	$\beta_{s0} \pm \Delta_{dof}\beta_{s0}$
Rotor pole arc, β_r	$\beta_{r0} \pm \Delta_{dof}\beta_{r0}$
Bore Radius, R_g	$R_{g0} \pm \Delta_{dof}R_{g0}$
Rotor Yoke Radius, R_{ry}	$R_{ry0} \pm \Delta_{dof}R_{ry0}$

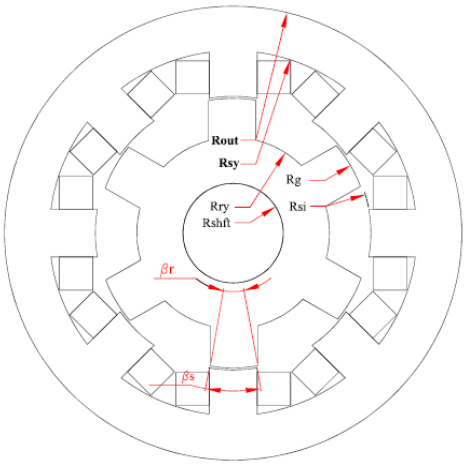

Figure 1. Cross-section of a 8/6 SRM and the associated DoFs.

Such a result would involve a change in the size of the cooling circuit which is undesirable. The problem was solved by means of adjusting the slot area. In order to control the slot area, the inner stator yoke radius, R_{sy} is considered as an additional DoF which is adjusted at each evaluation as such. If $\Delta\beta_s$, $\Delta\beta_r$, and ΔR_g are the differences between the values of the considered design variables at kth evaluation and their initial values, the change in R_{sy} can be computed using (2) where, θ=sin((β_s+$\Delta\beta_s$,)/2), R_{si}=R_g+ΔR_g+g and N_S is the number of stator poles. The additional DoF δk_f depends on the optimisation scenario. The tendency of the optimization routine is to narrow the stator yoke thickness. From the electromagnetic point of view reducing the yoke thickness down to a lower boundary (20% to 40% higher than half of the stator pole width) would be acceptable. On the other hand such a reduction, often deteriorates the performances, from the noise production as well as from structural point of view. In order to eliminate this inconvenient the structural constraints are included in the optimisation.

b. Mode frequencies computation

The analytical model described in [20] is used for the characterization of the sound power level radiated by the SRM. Three mode shapes are considered (m=0, m=2, m=4) and their own natural mode frequencies are calculated using (3) and (4).

The mode frequency for the pulsating vibration mode, $m = 0$

$$f_{m0} = \frac{1}{2\pi R_m}\sqrt{\frac{E_s}{\rho\Delta}}, Hz \tag{3}$$

whereas for modes $m \geq 2$ the following formula is used

$$f_{(m\geq 2)} = \frac{f_{m0}\cdot i\cdot m}{\left[\frac{m^2+1}{m^2-1} + i^2\left(4m^2 + m^2\frac{\Delta_m}{\Delta} + 3\right)\right]} \tag{4}$$

where

$$\Delta R_{sy} = \left(\frac{N_S}{\pi}\cdot R_{si}\cdot\theta - R_{sy}\right)\left(1+\sqrt{1-\frac{R_{sy}^2 - R_{si}^2 - 2\cdot\frac{N_S}{\pi}\cdot\theta\cdot\left(R_{sy}+R_{si}\right) - \frac{N_S}{\pi}\cdot A_0 - \delta k_f}{\left(N_S/\pi\cdot R_{si}\cdot\theta - R_{sy}\right)}}\right) \tag{2}$$

$$i = \frac{1}{2\sqrt{3}}\frac{h_{ys}}{R_m}$$

$$\Delta = 1 + \frac{W_t}{W_{sy}}$$

$$W_t = W_{sp} + W_{Cu} + W_i$$

$$\Delta_m = 1 + \frac{91N_s\cdot A_{sp}\cdot h_{ps}^3\cdot W_t}{R_m\cdot L_{stk}\cdot h_{ys}^3\cdot W_{sp}}\left[\frac{1}{3} + \left(\frac{h_{ys}}{2h_{ps}}\right) + \left(\frac{h_{ys}}{2h_{ps}}\right)^2\right]$$

where m, f_m are the circumferential mode number and mode frequency, E_S, ρ_S are the modulus of elasticity and density of stator material, Δ, Δ_m are the mass addition factors of displacement and rotation, W_{sp}, W_{sy}, are the weight of stator poles and stator yoke, W_{Cu}, W_i are the weight of the winding and insulation, R_m, h_{sy} are the mean radius and thickness of the stator yoke, N_s, h_{ps} are the number of stator poles and stator pole height and L_{stk}, A_{sp} are the stack length and cross-sectional area of each stator pole.

IV. RESULTS AND IMPLEMENTATION

A. B-H curve extrapolation

It is well known that SRM operates in higher saturation conditions as compared to classical AC machines [1]. On the other hand standard closed circuit magnetic induction measurement for electrical steel sheet is usually limited to 2T [17]. In such conditions the method used for extrapolation of BH curve, plays an essential role. In Fig. 2, the torque variation with position at constant phase current is shown for a B-H curve (M-19) in 2 cases:

Assuming that the maximum available value for magnetic flux density is 1.9T with the appropriate value for magnetic field intensity (after that the BH curve is linearly extrapolated) [18].

Assuming that the maximum available value obtained based on measured data for B is 2.3T.

In [17] multiple methods for the extrapolation of the BH curve of ferromagnetic materials used in the magnetic cores subject to high fields are compared. Further, the so-called, Law of Approach to Saturation Extrapolation (LAS) procedure is used for the extrapolation of the BH curve for the electromagnetic steel sheet M800-60A for which a value of B_{max}= 1.9T is available On the manufacturer website the $J(H)$ dependence is provided. Using Kennelly [18] field equation $B(H)$ dependence is obtained and the last 2 pair of points (B_{n-1}, H_{n-1}) and (B_n, H_n) are used for the computation of the magnetic flux density B_{high} which corresponds to an imposed field value, H_{high} as shown below:

$$B_{high} = \mu_0 \left(H_{high} + \frac{H_n^3}{b} \frac{(\mu_n - 1)}{2} \left(1 - \frac{b}{H_{high}^2} \right) \right) \quad (5)$$

where

$$b = \frac{\mu_0 H_n^3 \frac{(\mu_n - 1)}{2}}{B_n - \mu_0 H_n + \mu_0 H_n \frac{(\mu_n - 1)}{2}} \quad \text{with } \mu_n = \frac{1}{\mu_0} \frac{dB}{dH}$$

B. Results

The proposed optimal design methodology is assessed starting from the initial designs SRMPRE-S and SRMPPT for different optimisation scenarios explained hereafter.

Constant Current Density: First case considered is when only the 4 DoFs defined in Table 1 are taken into consideration for the SRMPRE-S. The yoke thickness is not modified in order not to deteriorate NVH characteristics.

Also the density of the current is kept constant at every evaluation (the electric circuit of the FE model has a current source which is characterised by its density, J_i). Allowing a maximum of 20% variation of each DoF (Δ_{dof}=0.2) the tendency of the minimization routine is to increase the slot area therefore increasing the required phase current as its density is kept constant.

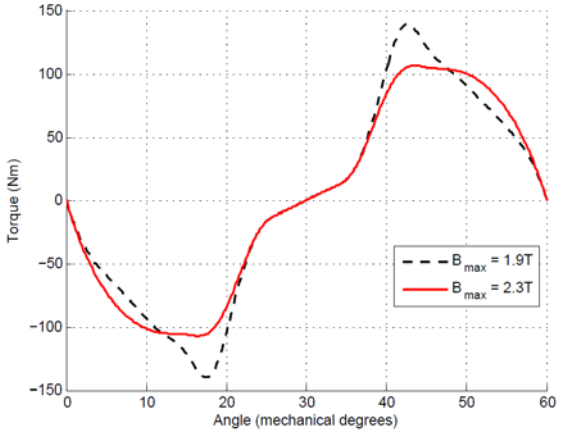

Figure 2. Torque obtained when different las points are assumed for the BH curve.

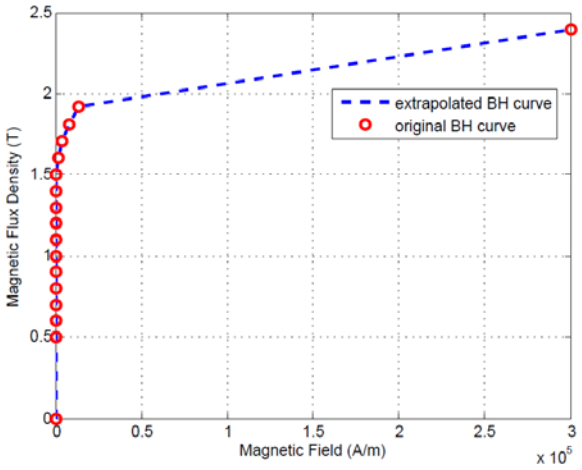

Figure 3. Original and extrapolated BH curve (using Law of Approach to Saturation Extrapolation method).

Despite the augmentation of the torque density as well as weight reduction (average torque increased by 13.3% and weight reduced by 8.75%) this solution is not attractive, especially for automotive applications, where the maximum current that can be provided by the battery is limited.

Constant slot area (J_i constant) - Scenario 1 The algorithm flow of the implemented routine is shown in Fig. 4. At each evaluation the current density is maintained constant by imposing a constant current in the electrical circuit of the FE model and automatically adjusting ΔR_{sy} such that, the slot area remains constant. For the computation of ΔR_{sy}, the condition ΔA=0 is applied, where $\Delta A = A_k - A_0$ is the change in the slot area, with A_k, A_0 the slot area at evaluation k and initial evaluation, respectively (2). Over each evaluation, the FE analysis is conducted three times. First analysis is executed for determining the value of magnetic flux density (in aligned position). If $B_{FE}<B_{max}$ then the next two FE analyses are conducted as follows: in each analysis the coenergy is computed and ultimately the approximated average torque by dividing the change in coenergy in respect to the change in position. Firstly, the value of coenergy is evaluated at position, $\theta_0=(\beta_s+\beta_r)/2$ which is the angle when poles start to overlap. Further, the coenergy is computed at position, $\theta_1=\theta_0+\theta_s$ where θ_s is the stroke angle and is given by $\theta_s=360^\circ/(N_{ph}N_R)$ with N_{ph} the number of phases. It must be stated that the maximum value allowed for magnetic flux density is matched with the maximum value provided with the B-H curve (see Section IV-A) and a maximum change of 30% is allowed for the DoFs. As for the current density a maximum value of J_i=25A/mm^2 was imposed.

The maximum value for the second mode shape frequency was chosen f_{m2}=800Hz for the SRMPRE-S. The reasoning in choosing this value is that the maximum speed (in rpm) that the SRM has to turn with, corresponding to 120kph cruise speed of the vehicle, i.e. 7200rpm. On the other hand, for an 8/6 SRM the fundamental of the excitation frequency of 800Hz corresponds to a speed of 8000rpm, thus the resonance with the corresponding second mode shape frequency is avoided.

However for the SRMPPT first tests were conducted considering f_{m2}=1000Hz as the maximum speed of the machine, provided in the datasheet is 10000rpm and eventually, just from a comparative point of view, allowing a lowering of the second mode shape frequency down to 800Hz.

Constant Current Density and higher number of turns - Scenario 2 Starting from the dependence of the current density with the number of turns ($J_i=N_p\ i/A_{coil}$), the variation of ΔR_{sy} can be calculated so as to ensure constant current density in the coil.

Therefore in (2), if $\delta k_f = A_0 \Delta N_p / N_p$, then the slot area is automatically adjusted such that the current density is kept the same as in the initial evaluation. In previous equation, the number of turns is denoted with N_p and the change in the number of turns at evaluation k as compared to initial value is denoted as ΔN_p.

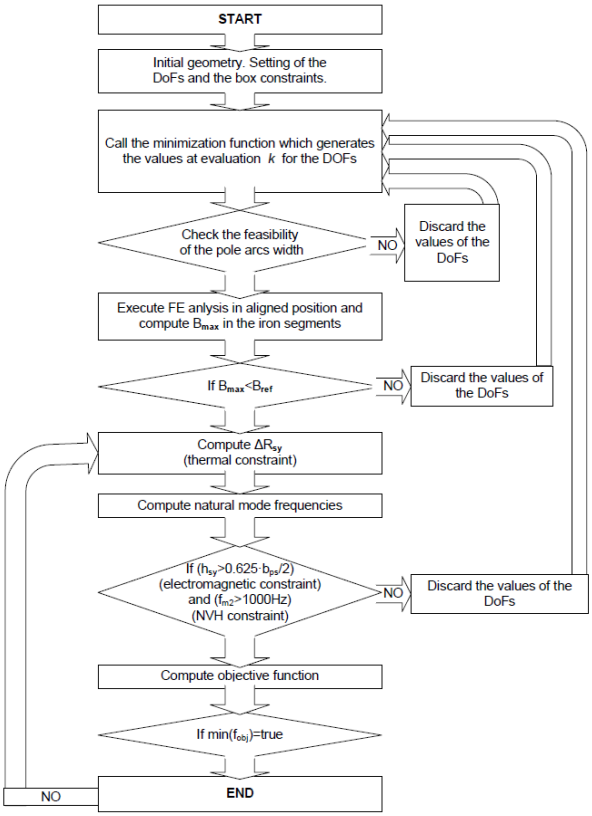

Figure 4. Algorithm flow of the optimisation routine.

As expected, increasing the number of turns yields higher torque (Fig. 5 and Fig. 6). Moreover, it clearly can be noticed the tendency of the minimization routine to reduce the stator yoke thickness and to fill the remaining space with conductors. This scenario proceeds until the structural constraints intervenes and halts the loop (when the second mode frequency decreases more than the set limit).

For both considered machines, optimizations are conducted for maximizing the torque density. The gain in torque (at rated current) is shown in Fig. 5 whereas in Table 2 are provided the values of the DoFs and constraints for the SRMPRE-S. Finally, in Fig. 6 the gain in torque is provided for the SRMPPT along with the values for DoFs and constraints in Table 3. Moreover the change in the analytically computed mode shape frequencies is provided as well as the initial (ID) and the optimized (OD) geometries (Table 2 and Table 3).

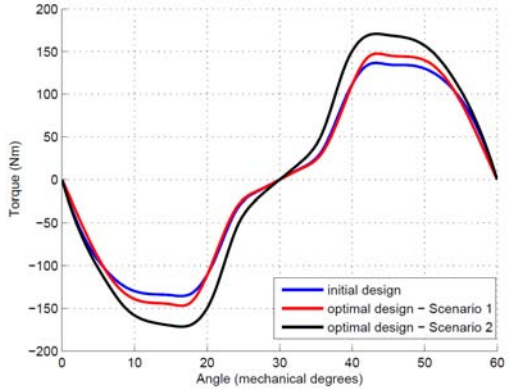

Figure 5. Torque obtained at rated current for initial and optimal – Scenario 1 and 2 for SRMPRE-S.

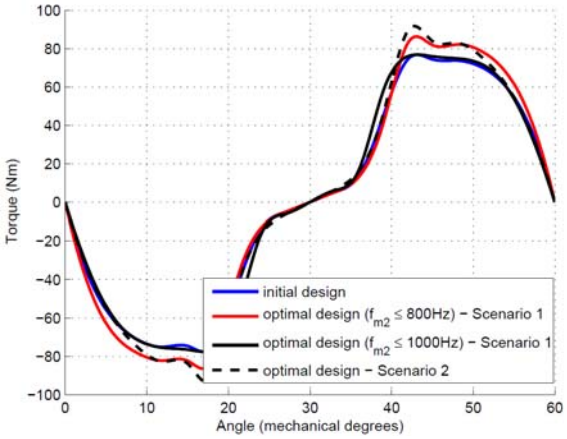

Figure 6. Torque obtained at rated current for initial and optimal – Scenario 1 and 2 for SRMPPT.

TABLE 2. The DOFs and the constraints associated with each optimization scenario for SRMPRE-S.

Quantities	**SRMPRE-S**		
	ID	OD-S1	OD-S2
$f_{m(=0)}$, [Hz]	5657	5518	5368
$f_{m(=2)}$, [Hz]	900	815	809
$f_{m(=4)}$, [Hz]	3259	2940	2415
β_s, [°]	20.12	18.31	20.14
β_r, [°]	20.35	21.43	20.74
R_g, [mm]	87	93.2	86.5
R_{ry}, [mm]	55.7	59.3	56.1
W_{FE}, [kg]	49.69	49.36	47.84
R_{out}, [mm]	144	144	144
L_{stk}, [mm]	121.8	121.8	121.8
N_p, (turns)	24	24	30
T_{av}, [Nm]	134.74	145.48	169.17
Gain torque, [%]	-	7.97	25.55
B_{max}, [T]	2.12	2.29	2.27
J_i, [A/mm^2]	24.98	24.98	24.98

TABLE 3. The DOFs and the constraints associated with each optimization scenario for SRMPPT.

Quantities	**SRMPPT**			
	ID	OD-S1		OD-S2
$f_{m(=2)\ max}$ [Hz]		800	1000	1000
$f_{m(=0)}$, [Hz]	7256	6693	7142	7096
$f_{m(=2)}$, [Hz]	1076	800.5	1000.1	1000.1
$f_{m(=4)}$, [Hz]	4056	2986	3835	3536
β_s, [°]	20	19.76	19.79	19.33
β_r, [°]	20	19.82	21.32	19.68
R_g, [mm]	61.2	67.2	62.9	58.3
R_{ry}, [mm]	43.2	45.7	41.2	44.9
W_{FE}, [kg]	36.49	36.24	36.47	35.07
R_{out}, [mm]	102.5	102.5	102.5	102.5
L_{stk}, [mm]	185	185	185	185
N_p, (turns)	50	50	50	59
T_{av}, [Nm]	74.43	82.91	75.96	89.37
Gain torque, [%]		11.39	2.05	20.07
B_{max}, [T]	1.22	1.56	1.20	1.55
J_i, [A/mm^2]	16.95	16.95	16.95	16.95

V. CONCLUSIONS

In this paper a multiphysics design optimisation routine was proposed and the performances were assessed using two different designs of Switched Reluctance Machines, by means of Time-Stepped Finite Element Models. The differences occurred regarding the methods used to obtain the input data for the FE Models. In first case (based on some imposed specifications), a SRM is pre-sized and then optimized in order to meet the requirements of an electric motor used in HEV/EV propulsion, taking into account some constraints, whereas in the second case an existing SRM is used and attempts were made to increase the torque density by means of using a multiphysics optimisation routine. It was shown that considering three fundamental types of constraints (electromagnetic, thermal and structural) the average torque can be increased (rather significant in some cases) without major changes of the rest of the considered quantities. The constraints embedded in each optimization scenario were chosen so as to ensure that neither the machine case, nor the cooling circuit is modified.

ACKNOWLEDGEMENT

This research was financially supported in the frame of the project "Doctoral studies in engineering sciences with purpose to develop knowledge-based society - SIDOC (abbreviation from Romanian)", Contract POSDRU/88/1.5/S/60078. The authors equally acknowledged the WBI funding through the program "Wallonie-Bruxelles International, bursaries for excellence."

REFERENCES

[1] T. Miller, Switched Reluctance Motors and their Control, Oxford, Ed. Magna Physics Publishing, 1993.

[2] A.-C. Pop, V. Petrus, J. Gyselinck, C. Martis, and V. Iancu, “Comparative study of different torque sharing functions for losses minimization in switched reluctance motors used in electric vehicles propulsion,” *accepted for* 13th International Conference on Optimization of Electrical and Electronic Equipment, 2012.

[3] P. Rafajdus, V. Hrabovcova, P. Hudak, and M. Franko, “Torque optimization of switched reluctance motor,” in Workshop on Variable Reluctance Electrical Machines, Technical University of Cluj-Napoca, September 2002.

[4] K. Bienkowski, J. Szczypior, B. Bucki, A. Biernat, and A. Rogalski, “Influence of geometrical parameters of switched reluctance motor on electromagnetic torque,” in Berichte Und Informationen Hochschule fur Technik und Wirtschaft, Dresten, January 2002.

[5] W. Aljaism, N. M., and J. Rizk, “Torque optimization for srm by changing stator and yoke geometry,” in International Conference on Communication, Computer & Power, February 2007.

[6] M. Moalem, C.-M. Ong, and E. Unnewehr, “Effect of rotor profiles on the torque of a switched-reluctance motor,” IEEE Transactions on Industry Applications, vol. 28, no. 2, March/April 1992.

[7] S. M., N. P., and V. M., “Magnetic analysis and comparison of switched reluctance motor with different stator pole shapes using 3D Finite Element Method,” in ICGST-ACSE Journal, vol. 9, 2009.

[8] M. N. Anwar, I. Husain, and A. V. Radun, “A comprehensive design methodology for switched reluctance machines,” IEEE Transaction on Industry Applications, vol. 37, no. 6, November/December 2001.

[9] R. Hodkinson and J. Fenton, Lightweight Electric/Hybrid Vehicle Design, Oxford, Ed. Magna Physics Publishing, 2001.

[10] R. Krishnan, Switched Reluctance Motor Drives Modelling, Simulation, Analysis, Design and Applications. CRC Press, 2001.

[11] V. Petrus, A. Pop, C. Martis, J. Gyselinck, and V. Iancu, “Design and comparison of different switched reluctance machine topologies for electric vehicle propulsion,” in XIX International Conference on Electrical Machines (ICEM), 2010.

[12] G. Henneberger and I. A. Viorel, Variable Reluctance Electrical Machines. Shaker Verlag, Aachen, 2001.

[13] T. Uematsu and R. Wallace, “Design of a 100 kw switched reluctance motor for electric vehicle propulsion,” in Tenth Annual Applied Power Electronics Conference and Exposition, 1995.

[14] S. Faid and S. B. P. Debal, “Development of a switched reluctance motor for automotive traction applications,” in The 25th World Battery, Hybrid and Fuel Cell Electric Vehicle Symposium & Exhibition, Shenzhen, China, November 2010.

[15] P. J. Lawrenson, J. M. Stephenson, P. T. BIenkinsop, J. Corda, and N. N. Fulton, “Variable-speed switched reluctance motors,” IEEE Proceedings, vol. 127, no. 4, July 1980

[16] M. N. Anwar and H. I., “Radial force calculation and acoustic noise prediction in switched reluctance machines,” IEEE Transactions on Industry Applications, vol. 36, no. 6, November/December 2000.

[17] A. E. Umenei, Y. Melikhov, and D. C. Jiles, “Models for extrapolation of magnetization data on magnetic cores to high fields,” accepted for publication in IEEE Transactions on Magnetics, 2011.

[18] D. Meeker, Finite Element Method Magnetics: OctaveFEMM, March 2009.

[19] C. P. Steinmetz and P. L. Alger, Lectures on Electrical Engineering, D. Publications, Ed., 1971, vol. 3.

[20] J. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, “Convergence properties of the Nelder-Mead simplex method in low dimensions," SIAM Journal of Optimization, vol. 9, no. 1, pp. 112-147, 1998.

Analytical Analysis of the Tubular Transverse Flux Reluctance Motor

Dan-Cristian POPA, Vasile-Ioan GLIGA, Loránd SZABÓ, Vasile IANCU

Electrical Machines and Drives Department
Technical University of Cluj-Napoca, Faculty of Electrical Engineering
400750 Cluj-Napoca, P.O. Box 358, Romania, e-mail: Dan.Cristian.Popa@mae.utcluj.ro

***Abstract* – The paper deals with a new type of tubular electrical machine with modular construction. It can develop high force at short strokes. The structure is of transverse flux machines class and it operates based on the variable reluctance principle. The starting point of the modular tubular motor construction is a linear transverse flux reluctance machine. The design algorithm of the proposed machine is described in details, as well as an analytic approach based on the equivalent magnetic circuit to validate the correct sizing of the machine. An example is given, and the values of the flux densities obtained in different parts of the machine from analytical analysis are in good accordance with the starting design data.**

***Keywords:* tubular motor, transverse flux machine, variable reluctance machine, magnetic equivalent circuit.**

I. INTRODUCTION

The most researches on the transverse flux machine, introduced at the beginning of the 80's in the last century, were focused on the rotary variants [1]. The achievements in the field of the linear transverse flux machines were considerably less significant. In this paper a new type of variable reluctance tubular machine is proposed. In general terms, from the idea standing on the basics of this machine both rotating and tubular machines can be obtained.

The machine in study is originated from the linear transverse flux reluctance machine in modular construction. A group of researchers from the Technical University of Cluj-Napoca have proposed such a structure with hybrid excitation, which resulted as a combination between the rotary variant of a transverse flux machine with passive rotor, as one of the already conventional types of transverse flux machines, and permanent magnets and a Sawyer motor [1].

A simpler variant from the constructive point of view, with similar performances as the structure described above, was obtained by removing the permanent magnets from the modules of the mobile armature, having only electromagnetic excitation.

In order to increase the developed tangential force, the surface of the teeth was enlarged [4] so that the module had the same section in all its parts.

In this case, the stator of the machine is the passive armature and the mover, which carries the coils, represents the inductor, having a minimum number of three modules. The main shortcoming of a linear machine is given by the existence of a big normal force, of about ten times greater than the tangential one. By unfolding the structure without permanent magnets and enlarged surface of the teeth, Fig. 1b, after the direction of movement, a tubular variant, also in modular construction, is obtained [3].

II. THE STRUCTURE OF THE TUBULAR TRANSVERSE FLUX RELUCTANCE MACHINE

The inductor, usually the stator, has a modular construction, Fig. 1. The minimum number of phases N required in order to obtain a continuous movement is three [3, 4]. The iron core of a phase is built of m magnetic pieces, called modules, alternating with $m-1$ non-magnetic pieces, named spacers, Fig. 1. Each phase has an independent winding. The stator pieces have the same construction like the stator of a SRM. Each module can be made either of classical steel sheets or of soft magnetic composite material. Each module has Z poles and slots. The required distance between them is assured by the non-magnetic pieces, with different axial length than the stator spacers.

A phase's axial tooth and slot form together the tooth pitch τ. As in the case of the linear machine, the positioning step is given by the tooth pitch and the number of modules.

In the case of the linear transverse flux reluctance machine, in order to work properly the modules have to be shifted one from each other by $k\tau + s_m + \tau/N, \quad k \in \aleph$, where τ and N have the same significances like those above mentioned for the tubular machine. This condition is applied also to the motor proposed here. But, unlike the linear machine where the modules were placed at the construction in an aluminum case in such way that this shifting between the modules was provided, at the tubular motor the shifting is created by using non magnetic spacers, Fig. 1 [2].

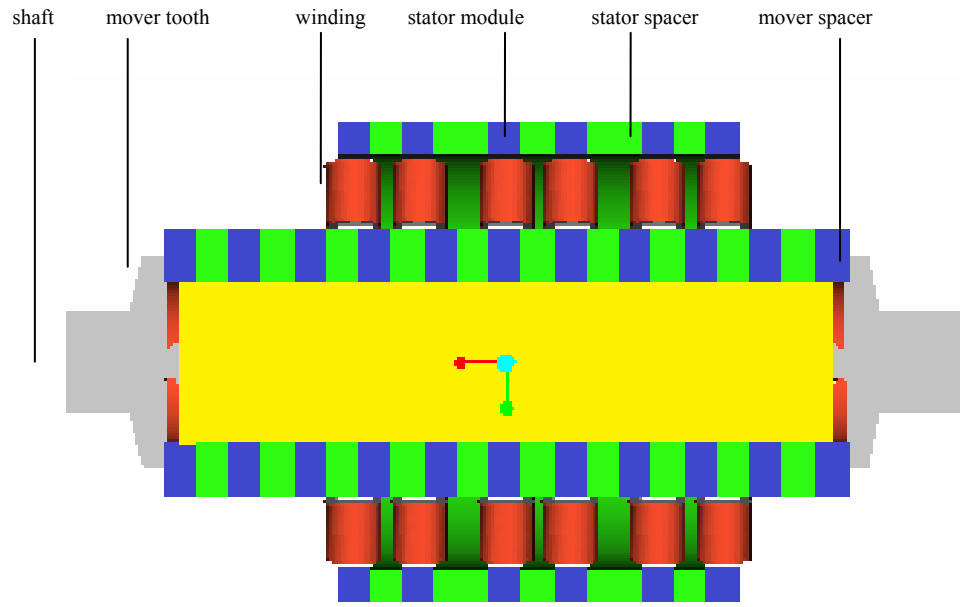

Figure 1. Longitudinal section through tubular transverse flux reluctance machine.

The induced armature, the mover, is passive in the case of the tubular machine, too, and the toothed structure must be obtained as well, but it will be in fact a cylinder. The construction with magnetic and non-magnetic pieces can be used as in the stator. The form of the pieces is much simpler than for the other armature, just a magnetic or non-magnetic cylinder, with a place in inner part for assembling the shaft, made also of non-magnetic material, Fig. 1 [2]. The advantages given by this construction are, as in the case of the stator, a lower weight and cost.

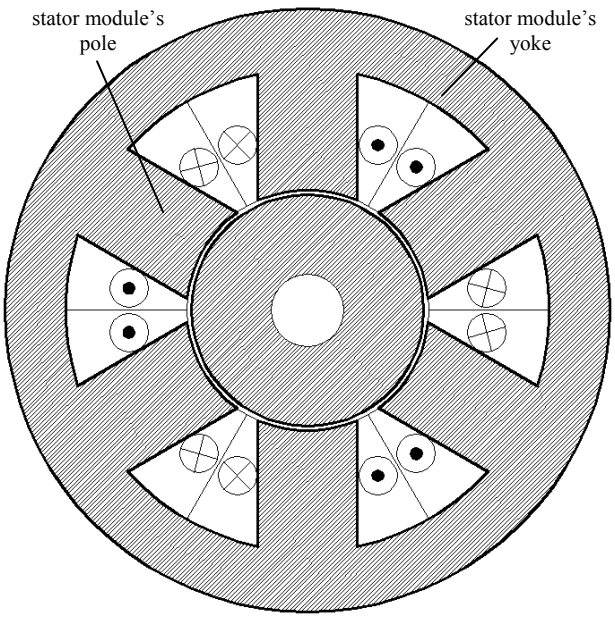

Figure 2. Cross section through the transverse flux tubular machine.

The disposal of the windings of the tubular machine is shown in a cross section of the machine, Fig. 2. The variant with concentrated windings around each tooth of the stator, connected in series, is presented. Another possible solution would be with coils wound around the yoke.

III. DESIGN ALGORITHM

The most important characteristic of the linear transverse flux reluctance machine is the developed tangential force. The normal force is about ten times bigger than the tangential one, and this is one of the major shortcomings of all linear machines [5]. As stated before, the existence of only the traction force due to the compensation of all the attraction forces is one of the most important advantages of the tubular variant. The force is given, like in the case of the linear structure, by a single module with energized winding. The main dimensions and excitation magneto-motive force (mmf) strongly depend on the required traction force. This motor has the particularity that the iron core of the two armatures is not homogenous. Considering that this introduces major difficulties from the first steps of the design stages, one can start from the hypothesis that the whole structure is made of magnetic material, the error being relatively small.

The traction force of any linear motor can be calculated analytically or by finite element analysis. In this paper, the basic aspects of an analytic approach shall be covered. The principle used to compute the developed forces at any linear variable reluctance machine is the variation of the magnetic energy in the air-gap versus the linear displacement [6]. The expression of the magnetic energy, where the elemental volume is function of mover's position, is (1).

$$W_m = \frac{1}{2}\cdot\frac{B_g^2}{\mu_0}\cdot\int dv = \frac{1}{2}\cdot\frac{B_g^2}{\mu_0}\cdot A_p\cdot g = \frac{1}{2}\cdot\frac{B_g^2}{\mu_0}\cdot R\cdot\alpha\cdot g\cdot(t_m - x) \quad (1)$$

$$f_T = \frac{\partial W_m}{\partial x} = \frac{1}{2}\cdot\frac{B_g^2}{\mu_0}\cdot g\cdot(-R\cdot\alpha) = -\frac{1}{2}\cdot\mu_0\cdot F^2\cdot\frac{R\cdot\alpha}{g} \quad (2)$$

$$F = \frac{1}{\mu_0}\cdot g\cdot B_g \quad (3)$$

The traction force f (2) and the coil mmf F (3) are expressed above, where the notations are: A_p – common armatures area, g – air-gap length, R – stator inner radius in the air-gap, α – stator pole angle, x – axial coordinate B_g – peak air-gap flux density value in aligned position.

The force f (3), is constant and does not depend on the axial length of the armatures, but on the square of the coil's mmf, pole circumferential length in air-gap ($R\alpha$) and air-gap length g. Considering the structure of the tubular motor proposed here, the traction force developed by an energized phase is:

$$f_{ph} = \mu_0\cdot F^2\cdot\frac{R\cdot\alpha}{g} \quad (4)$$

Starting from these relationships, a connection between the force and the magnetic and geometric dimension of the machine can be established:

$$R = \frac{\mu_0}{4}\frac{K_C K_S}{r}\frac{f}{C_r\cdot B_g^2\cdot g\cdot\alpha\cdot m\cdot Z} \quad (5)$$

where K_C, K_S are the Carter's and saturation coefficient, r is the ratio between the common axial length of the stator and mover pole and the polar pitch, C_r is the air-gap equivalent reluctance coefficient, Z is the number of poles of a module.

One must take into account that K_C and C_r coefficients are function of the air-gap length to mover pole pitch g/τ and mover pole axial length to mover pole pitch, lp/τ ratios, and also that the value of α is at designer's choice.

Considering that the area of a slot is given by

$$A_s = \frac{2g\cdot B_g}{\mu_0 J_1\cdot K_{fill}} \quad (6)$$

and taking into account some simplifying geometric hypothesis, one can obtain the height of the pole h_p:

$$h_p = \frac{\sqrt{2\cdot A_s\cdot\sin\alpha + (\frac{2\pi}{Z}-\alpha)\cdot\frac{(1+\cos\alpha)}{2}\cdot[R^2\cdot(\frac{2\pi}{Z}-\alpha)+2\cdot A_s]} - (\frac{2\pi}{Z}-\alpha)\cdot R\cdot\cos\frac{\alpha}{2}}{(\frac{2\pi}{Z}-\alpha)\cdot\frac{(1+\cos\alpha)}{2}} \quad (7)$$

All the other geometric dimensions result easily. The exterior radius of the stator is computed considering the flux that closes through the yoke at a typical SRM [7].

$$R_e = R\quad(1+0{,}8\quad\frac{\alpha}{2}) + h_p\quad\cos\frac{\alpha}{2} \quad (8)$$

The proposed algorithm was applied in order to obtain a motor developing a maximum tangential force of 500 N. The chosen values are: peak air-gap flux density B_g = 1.38 T, current density $J = 5\cdot10^6$ A/m^2, air-gap length g = 0.5 mm, slot fill factor K_{fill} = 0.4, saturation coefficient K_S = 1.4, Carter's factor K_C = 1.6, air-gap equivalent reluctance coefficient C_{rc} = 1.3, stator pole axial length per pole pitch r = 1/2. For the proposed machine, Z = 6 poles were considered, and the number m of modules is 2. The angle α of a pole was considered to be of 50^o.

The necessary area of the stator slot is (6) A_s = 5.5 cm^2 and the required mmf per coil is (3) F = 1650 Aturns. In this case the height of the pole was computed (7), h_p = 50 mm, the exterior radius being (8) R_{ex} = 115 mm. the motor is supplied with a voltage U = 120 V and has a rated current I = 37.5 A, the number of turns being 44.

IV. ANALYTICAL ANALYSIS USING THE MAGNETIC EQUIVALENT CIRCUIT

In order to check the validity of the proposed design algorithm, an equivalent magnetic circuit for the proposed structure was built up, Fig. 3 [8]. The circuit is for three poles, the yoke parts and the slots between them and the corresponding part of the mover. Besides the mmf of a coil F, it is known, using the determined geometric dimensions, the reluctance of the yoke R_y, of the pole R_p, of the air-gap R_g and of the mover R_m. The leakage reluctances $R_{\sigma p}$ and $R_{\sigma s}$ are due to the air between two neighboured coils and to the air-gap.

The problem is to find out the fluxes through each branch of the circuit. It is solved by applying the Kirchhoff's laws at the obtained circuit [9]. One must underline that the first equation can be written correctly only for the middle branch. The solution of the obtained equation system results easily using modern programs.

Consequently, one can compute the percentage value of each flux from the greatest one in the circuit which is for this motor, the one in the poles. The flux density value in the poles, yoke, mover and air-gap is then computed and then compared with the imposed design data.

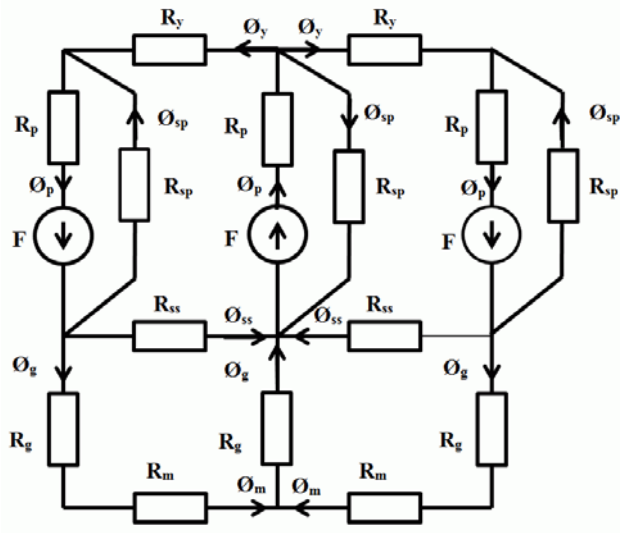

Figure 3. Magnetic equivalent circuit of the tubular machine.

The analytic computation was done for a shifting of half of mover piece from the stator one. As it results from the magnetic circuit, the biggest value of the flux is in the poles. Considering this as the reference value, the percentage value of 83.89% is obtained for the flux in the yoke, 72.7% for the air-gap flux, and 11.2%, respectively 16.1% for the two leakage fluxes. The mean value of the air-gap flux density is of 1.27 T, in very good accordance with the chosen one in the design

V. CONCLUSIONS

This paper deals with a new type of tubular machine, belonging to the transverse flux machines class and operating based on the variable reluctance principle. The design algorithm which is proposed for this structure was verified using an analytical method based on the equivalent magnetic circuit of the machine. The machine is suitable for applications requiring precise positioning step or high forces at low speed, with reduce strokes.

ACKNOWLEDGMENTS

This paper was supported by the project "Development and support of multidisciplinary postdoctoral programmes in major technical areas of national strategy of Research - Development - Innovation" 4D-POSTDOC, contract no. POSDRU/89/1.5/S/52603, project co-funded by the European Social Fund through Sectorial Operational Programme Human Resources Development 2007-2013."

REFERENCES

[1] D.C. Popa, "Theoretic and Experimental Study of the Linear Transverse Flux Reluctance Motor" (in Romanian), Ph.D. thesis, Technical University of Cluj-Napoca (2008).

[2] D.C. Popa, V.I. Gliga, V. Iancu, "*Tubular Transverse Flux Reluctance Motor in Modular Construction",* Annals of the University of Craiova, Serie: Electrical Engineering, Year 34, no. 34, 2010, Vol. II, pp. 79-83 (2010).

[3] D.C. Popa, V. Iancu, L. Szabó, "*Linear Transverse Flux Motor for Conveyors"*, Proceedings of the 6th International Symposium on Linear Drives for Industrial Application (LDIA '2007), Lille (France), paper 188, on CD (2007).

[4] J.F. Llibre, N. Martinez, B. Nogarede, P. Leprince – Linear tubular switched reluctance motor for heart assistance circulatory: Analytical and finite element modeling, 10th International Workshop on Electronics, Control, Measurement and Signals (ECMS), pp. 1-6 (2011).

[5] I. Boldea., "Linear Motion Electromagnetic Devices", Taylor & Francis, U.S.A. (2001).

[6] Larisa Strete, I.A. Viorel, "*Modular Tubular Switched Reluctance Motor"*, Proceedings of PCIM '08, vol. Intelligent Motion, Nürnberg (Germany), on CD (2008).

[7] G. Henneberger, I.-A. Viorel, "Variable Reluctance Electrical Machines", Shaker Verlag, Aachen, 2001.

[8] M. Ruba., "Design and Study of a Modular Switched Reluctance Machine", Ph.D. Thesis, Technical University of Cluj-Napoca (2008).

[9] R. Krishnan, "Switched Reluctance Motor Drives - Modeling, Simulation, Analysis, Design, and Applications", Industrial Electronics Series, CRC Press, 2001.

Performance Improvement of DFE on CDMA channel

PORURAN Sivakumar[1], MARIMUTHU Rajaram[2],
[1] Department of Electronics and Communication Engineering,
SKP Engineering College Thiruvannmalai, India,
E-Mail: sivakumar.poruran@gmail.com

[2] Department of Electrical and Electronics Engineering,
Anna University of Technology Thirunelveli India,
E-Mail: rajaramgct@rediffmail.com

Abstract *– – During last few years, there has been a dramatical rise in commercial, interest in code division multiple access (CDMA) systems. A CDMA system provides protection against interference, multipath, fading and jamming. Because of this, it yields capacity improvements over other access schemes. There are several interference cancellation schemes for CDMA, which require information about all interfering active users or some channel parameters. In this paper, we present a decision feedback equalizer (DFE) for a CDMA system in an indoor wireless Rayleigh fading environment. This system only uses information about the desired user's spreading code and a training sequence. Bit error probability (BEP) analysis of the system shows significant capacity improvements. Simulation results are included.*

Keywords: *DEF, WCDMA, multipath, fading, bit error probability (BEP), ISI*

I. INTRODUCTION

In CDMA system, many users can share the same bandwidth. All the different users are assigned different spreading codes to spread their signals over a much wider bandwidth than their transmitted data bandwidth.

In a direct sequence spread spectrum (DS/SS) CDMA technique, the transmitted data stream is multiplied by a higher rate spreading code to spread the information's energy over a wider bandwidth. In past, spread spectrum techniques were implemented in military communication systems since they provided protection against interference, multipath, fading and jamming.

Commercial interest in mobile and portable communications has increased demands to allocate more bandwidth for radio communications, or use more efficient transmission schemes to provide services for wireless communications. One CDMA system designed and implemented by Qualcomm Incorporated [1] shows considerable improvement in capacity over other FDMA and TDMA systems, since it exploits multipath resolution, voice activity, and antenna sectorization as well as the spreading gain. The main capacity gain comes from the fact that the same bandwidth could be used in neighbouring cells of the cellular system, which is difficult in both FDMA and TDMA systems. Using 1.25 MHz bandwidth, a data rate of about 9600 bits per second and$E_b/N_o$of about 6 dB, the Qualcomm system could provide 120 voice channels in a given cell that uses 120º sectored antennas and voice activity detection.However, this capacity drops to 33 channels if an Omni directional antenna is used and voice activities of users are not considered [2].

CDMA systems using linear and nonlinear interference canceling techniques have shown even better performance if the receiver has knowledge of the spreading codes of all users, received powers of some of the interferers, or some channel parameters [3–5]. All spreading codes pertaining to a given cell in a cellular radio system are available at the receiver, but it is difficult to provide codes from other cells to the base station.

It is also difficult to provide any interfering codes to individual portable receivers. Security is also of a concern when a receiver possesses knowledge of the spreading codes of other users in a given cell. Further, it is also very tedious to estimate channel parameters for the fast varying channel of mobile communications. Considering the drawbacks of the assumptions of these proposed systems for CDMA, a new scheme using decision feedback equalizer (DFE) for CDMA systems is used.

The choice of a DFE receiver for CDMA was supported by previous research showing that a DFE receiver improves the performance in a near-end crosstalk (NEXT) environment of digital subscriber loops (DSL) [6]. The DFE combines the functions of RAKE reception [7] to exploit diversity resulting from multipath, and performs interference and intersymbol interference (ISI) cancellation. The DFE minimizes the effects of interference as well as ISI by trying to force zeros in the impulse responses of the interferers at the decision instants. The number of interferers that theoretically can be canceled increases with the signal's excess bandwidth. For a base band bandwidth that is S times the data symbol rate, upto 2S – 1 interferers can be canceled [8–9].

Two configurations of the DFE receiver have been suggested for CDMA [10-12]. One uses information about the desired user's spreading code, while the other assumes no knowledge of any spreading code. In both configurations, we do not assume any information about channel parameters or user's activities but a known training sequence of data symbols is assumed in most cases. This reduces the complexity of the CDMA system, and provides security for users. It do not want other systems to have access to their spreading code.

II. DECISION FEEDBACK EQUALIZER

The performance of the linear equalizer is limited in the case of severe ISI. The Decision Feedback Equalizer (DFE) is the improved equalizer according to the linear equalizer by the introduced nonlinearity. The DFE has a noise reduction according to the linear equalizer which suffers from noise enhancement. DFE is composed of two filters. That is feedforward filter and feedback filter. Both are implemented to work at, in general, symbol rate. The feedforward filter can be fractionally spaced and it is a transversal filter, gets the output of the receive filter and sampler as input. So, it is some way like the linear equalizer.

The feedback filter gets the past decisions of the transmitted symbols at the output of the decision device as input. The feedback filter uses the past decisions for its output, it is strictly causal. The block diagram of the DFE is given in Figure 1. The feedforward filter removes some of the ISI from the received signal, but leaves some of the post cursor ISI on the signal.The feedback filter estimates the residual ISI from the past decisions and subtracts it from the feedforward filter output. The DFE solution is better than linear equalizer with a low- complexity solution.

The low noise enhancement of the DFE arises by assuming no decision errors, the decision device removes all the noise present in the signal. So, the inputs of the feedback filter have no noise, so the outputs of the feedback filter have no noise. Also, the feed forward filter has the less complex problem of removing only precursor coefficients. This results in a better performance for the feed forward filter according to the linear equalizer. But, the assumption of correct symbol decisions at the output of the decision device may not work in practical cases, so error propagation occurs and the performance of the equalizer is degraded.

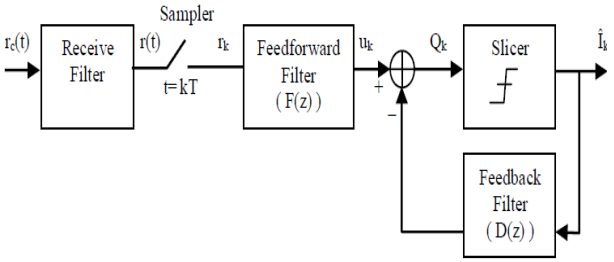

Figure 1. Decision feedback equalizer.

III. PROPOSED METHOD

A. Transmitter

Transmitter part of the CDMA system used in the paper is shown in figure 1. The transmitted signal of user k is a spread spectrum BPSK signal represented as:

$$S_k(t) = \sqrt{2}\,\mathrm{Re}\left(\sum_{\lambda} a_\lambda^k\, b_k(t-\lambda T_b - t_k)\right) e^{j2\pi f_c t} \tag{1}$$

Where:

aλ is the real-valued transmitted data Symbol ±1

b_k is the spreading sequence of user k

T_b is the symbol period

T_k is the delay of user k with respect to the desired user

f_C is the carrier frequency.

The spreading sequence of user k is defined as:

$$b_k(t) = \sum_{\lambda=0}^{l-1} b_\lambda^k\, p(t-\lambda T_c) \tag{2}$$

Where:

b_k is the spreading code

T_c is the chip period

l is the length of the spreading code in terms of chip periods

p(t) is the transmitted pulse shape.

Each user's transmitted signal is assumed to pass through a frequency selective Rayleigh fading channel. Additive white Gaussian Noise (AWGN), resulting from receiver thermal noise, is also considered in this system. The received chip-energy to noise ratio, Ec/No, which is also the receiver's input signal to thermal noise ratio, is assumed to be 6 dB.

With a spreading gain of 8, the corresponding E_b/N_o ratio is 15 dB. Here, No does not include interference from other users.

B. Receiver

Figure 2 shows the block diagram of receiver "0". The received signal, from N + 1 users, after demodulation is represented as:

$$d(t) = \sum_{k=0}^{N} d_k(t) + n(t) \tag{3}$$

Where:

$$d_k(t) = \sum_{\lambda} a_\lambda^k \sum_{m=0}^{M_k} C_{mk}\, b_k(t-\lambda T_b - t_k - t_{mk}) \tag{4}$$

M_K is the number of paths of user k

C_{mk} is the complex quantity which represents the amplitude and phase variation of the mth path of user k

t_{mk} is the reception time of the mth path of user k.

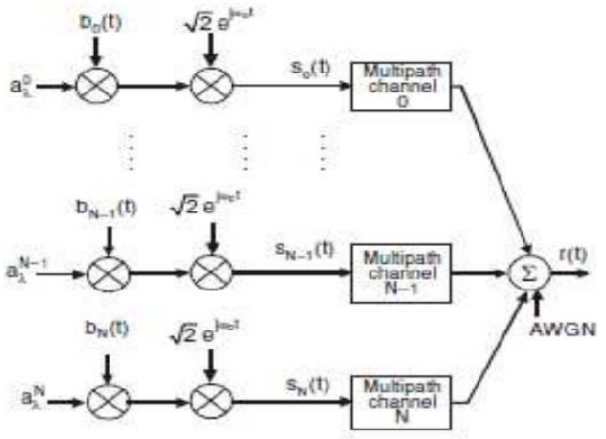

Figure 2. CDMA transmitter.

Two receiver configurations have been considered. In one configuration, the demodulated signal passes through a spreading sequence matched filter (SSMF) to the desired user before being received by the DFE. In the other receiver configuration, the SSMF is replaced by a low pass filter (LPF), which has a bandwidth equal to the spreading code bandwidth since the DFE is sampling at multiples of the chip rate of the spreading code.

The transmitted signal of each user is assumed to be received at the input of the DFE receiver through three different paths ($MK = 3$; $k = 0, 1 \ldots$). To determine the strength of each path, two types of channels were assumed. In one case, which is called as the constant variance model, all paths are assumed to have equal mean and variance, which represents a pessimistic type of channel.

The other channel model has the variance of a given path decreasing by $e^{-(L-1/3)}$, where L is the time, measured in the chip periods of the spreading sequence, with respect to the arrival time of the first path. This type of channel is called the exponential variance model. We have also assumed steady-state channels, which means that the channel coefficients do not change during the data transmission period of a user. The input to decision device is equal to

$$y_n = \sum_{l=0}^{N_Z} z^*_l \, g(nT_b - l\Delta) + \sum_{h=1}^{N_f} f_h^* \, a_{n\ h\ D}^{0} \qquad (5)$$

Where:

$N_z + 1$ is the number of forward taps

N_f is the number of feedback taps

Z_l is the setting of the lth tap of the forward filter

f_h is the setting of the hth tap of the feedback filter

$\Delta = T/M$, where M is an integer, and D is the delay assumed in the system.

In order to determine the optimum performance of the DFE receiver, the forward and feedback taps of the DFE are optimized to minimize the MSE,where the error is

$$e_n = y_n - a_{n-D}^{0} \qquad (6)$$

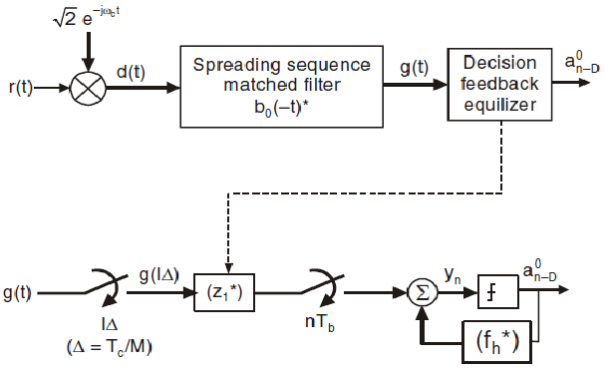

Figure 3. CDMA receiver.

The processing gain (PG), or ratio of spread bandwidth to information bandwidth, of a CDMA system is usually chosen in the order of 100. Since, we are using a DFE receiver for equalizing the effects of fading and interference and the forward filter of the DFE performs its sampling at multiples of the chip rate of the spreading code, then the PG should be moderate because higher sampling rate increases the complexity of the overall system. In this work, we have used a processing gain equal to 10^{-2}. If N'o is defined as the equivalent power spectral density of interference plus thermal noise, the signal- to-noise ratio per bit, of a conventional CDMA receiver is [1].

$$\frac{E_b}{N_o'} = \frac{PG}{N + N_o / E_c}. \qquad (7)$$

Where:

PG is the processing gain

N + 1 is the total number of users

Using MMSE criterion, the forward filter of the DFE receiver works on maximizing the desired user's energy and minimizing the effects of existing interference in the channel, AWGN, and ISI. The performance of the proposed system is evaluated by studying the MMSE of the DFE receiver.

IV. SIMULATION AND RESULT

The MMSE has been calculated in terms of the number of forward filter taps. The number of feedback taps has been fixed at 2, since the results showed that MMSE does not improve for a number of feedback taps greater than 2.

The DFE receiver was optimized to the best sampling phase and decision delay. To simplify the simulation of the DFE receiver, we have used fixed sampling phase and equalizer delay.

The Normalized Least-Mean-Squared (NLMS) algorithm was used to update the coefficients of the forward and feedback taps of the DFE.

In Figure 4, bit error probability of chip rate DFE with SSMF has been plotted versus the number of active users, k. Figure 5 shows the graph between bit error probability of DFE with LPF.

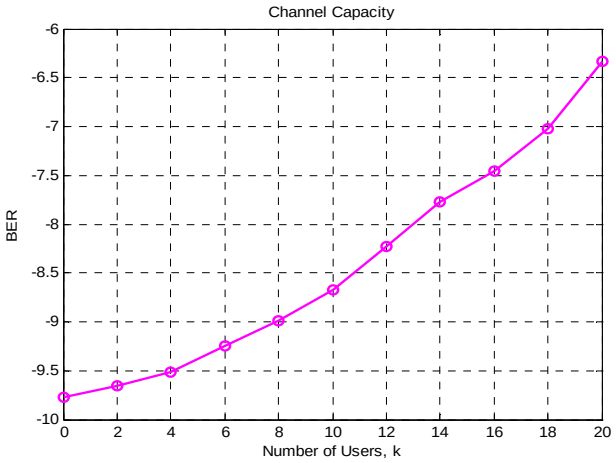

Figure 4. Bit error probability of DFE with SSMF.

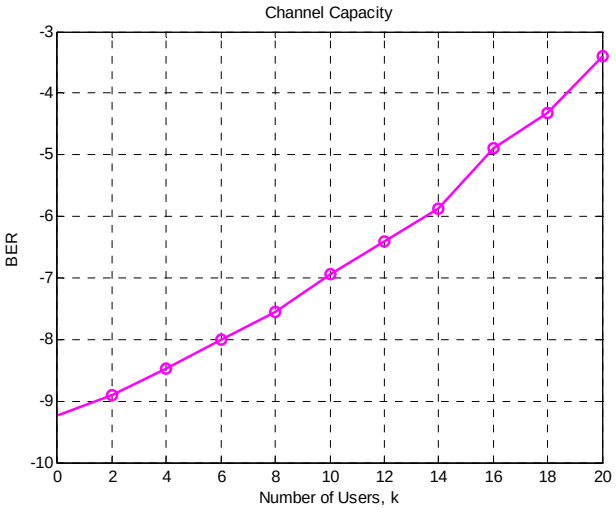

Figure 5. Bit error probability of DFE with LPF.

V. CONCLUSION

In this paper, DFE for WCDMA system has been investigated. The optimum performance and BEP results show the capabilities of this receiver in minimizing the effects of interference, multipath fading, and AWGN in a slow acting power control environment. Two receiver configurations have been considered: one that has knowledge of desired user-spreading code (SSMF receiver), and the other has no information about all spreading codes (LPF receiver).

Looking at the bit error probability (BEP) of this DFE receiver, figure 4 shows that eight users provide a good BER when using a SSMF at the input of the DFE. In calculating the BEP, one trial was performed for each channel and each user transmits 50,000 data bits. Figure 5 shows the BEP when the SSMF is replaced by a LPF. The results are worse in this case because the LPF receiver uses the same number of forward taps as the SSMF receiver and has no knowledge of the spreading code of the desired user but still provides good BEP for eight users. BEP in the order of 10^{-2} is acceptable in the system for voice transmissions. The BEP of the LPF case and the corresponding overall MSE could be improved by adding more forward taps. It was also noticed that the results of MSE and BEP are similar in providing the same conclusion in terms of the capability of the DFE.

REFERENCES

[1] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, Jr., and C. E. Wheatley 111, "On the capacity of a cellular CDMA system,'' IEEE Trans. Vehic. Technol., vol. 40, no. 2, pp. 303-312, May 1991.

[2] A. Salmasi and K. S. Gilhousen, "On the system design aspects of code division multiple access (CDMA) applied to digital cellular and personal communications networks," in IEEE 4Ist Vehic. Technol. Conf., 1991, pp. 57-62.

[3] R. Lupas and S. Verdu, "Near-far resistance of multiuser detectors in asynchronous channels," IEEE Trans. Commun., vol. 38, no. 4, pp. 49 C508, Apr. 1990.

[4] Z. Xie, R. T. Short, and C. K. Rush forth, "A family of suboptimum detectors for coherent multiuser communications," IEEE J. Select. Areas Commun., vol. 8, no. 4, pp. 683-690, May 1990.

[5] R. Kohno, H. Imai, M. Hatori, and S. Pasupathy, "Combination of an adaptive array antenna and a canceller of interference for direct sequence spread-spectrum multiple-access system,'' IEEE J. Select. Areas Commun., vol. 8, no. 4, pp. 675-682, May 1990.

[6] M. Abdulrahman and D. D. Falconer, "Cyclostationary crosstalk suppression by decision feedback equalization on digital subscriber loops,"IEEE J. Select. Areas Commun., vol. 10, no. 3, pp. 640-649, Apr. 1992.

[7] G. L. Turin, "Introduction to spread-spectrum antimultipath techniques and their application to urban digital radio," Proc. IEEE., vol. 68, no. 3, pp. 328-353, Mar. 1980.

[8] B. R. Petersen and D. D. Falconer, "Minimum mean square equalization in cyclostationary and stationary interference-Analysis and subscriber line calculations," IEEE J. Select. Areas Commun., vol. 9, no. 6, pp. 931-940, Aug. 1991.

[9] D. D. Falconer, M. Abdulrahman, N. W. K. Lo, B. R. Petersen, and A. U. H. Sheikh, "Advances in equalization and diversity for portable wireless systems," Digit. Signal Process. vol. 3, no. 3, pp. 148-162, Mar. 1993.

[10] M. Abdulrahman, D. D. Falconer, and A. U. H. Sheikh, "Equalization for interference cancellation in spread spectrum multiple access systems, "in IEEE 42nd Vehic. Technol. Conf., 1992, pp. 71-74.

[11] P. Sivakumar, Dr. M. Rajaram, "Performance of Frequency Domain Decision Feedback Equalization Using Sub Carrier Allocation," I.J. Information Technology and Computer Science, 2012, 1, pp 19-24.

[12] P. Sivakumar, Dr.M.Rajaram, "Performance Evaluation of Non Linear Equalization based on MLP for OFDM Power line Communication," I.J. Elec. Engg,Vol 4 no.8, 2011, pp 929-938.

Peer-to-Peer Overlay Network for On-demand Video Streaming

RAGAB Khaled

College of Computer Sciences and Information Technology,
King Faisal University, Hofuf, Saudi Arabia.
(on leave from Ain Shams University, Cairo, Egypt)
kabdultawab@kfu.edu.sa

Abstract *– Recently Peer-to-Peer (P2P) file downloading and streaming applications have attracted a large number of users on the Internet. The P2P video streaming systems are classified into two categories live and on-demand streaming systems. This paper surveys the current literatures on the existing P2P video streaming solutions. The author proposes a P2P video overlay network (P2P-VON) of peers for video on-demand (VoD) streaming. In P2P-VON, each peer is able to cache some video minutes associated with the current media being played. P2P-VON is organized into clusters. Each cluster contains peers with overlapped buffer windows where their playing points are located between lower and upper playing point limits. When a peer in the cluster moves its playing point within the limits, for example by performing a seek operation, it then can rapidly discover and fetch the required segments for the playback buffer from peers in the same cluster. Clustering scheme improves both discovery and fetching overheads. However, it needs cluster management overhead. To improve performance of the VoD on P2P-VON, a discover-to-exchange schedule algorithm has been proposed and investigated through simulation.*

Keywords: *Overlay Networks, Peer-to-Peer Computing, Video on-demand streaming.*

I. INTRODUCTION

The video streaming applications have recently fascinated a large number of users over the Internet. According to *AccuStream iMedia Research*, the number of video streams served increased 52% in 2008, and reaches 34 billion viewers. In addition, over the last ten years audiences have gained access 142.7 billions of pieces of video [11]. According to *comScore*, in July 2009 around 91 million viewers watched five billion online video on *YouTube*. Americans watched 558 million hours of online during that month [8]. Some forecasts predict that by 2015 the share of video content will exceed 90% of the global consumer traffic, as a result every second, one million minutes of video content will cross the Internet in 2015 [7]. Moreover, the amount of *VoD* traffic in 2015 will be equivalent to 3 billion DVDs per month. The most familiar solution for streaming video over the Internet is the client-server service model such as *YouTube*. In this model every demand is handled by a centralized server, requiring a powerful server and large bandwidth. Although server clusters can improve the processing ability, there is a limit on the bandwidth of a server. Because *VoD* streaming requires significant bandwidth, Client/Server solutions cannot support large number of demands over the Internet. For example, with 100Mb bandwidth and a 200Kbps video, the limit is only 500 concurrent demands, which is far from supporting *VoD* of hot events, such as the Football world cup or popular news video. The client/server model uses variants of technologies such as Content Delivery Network (*CDN*) to push video content from server to its clients through delivery servers [22]. Scalability is the major challenge for server based video streaming solutions as follows. Despite the fact that *Google* has its impressive *CDN*, on 23 November 2008, it relied on a third party *Akamai*, to stream a YouTube live concert to 700,000 concurrent viewers. Akamai technologies evolved out of an *MIT* research effort aimed at solving the flash crowd problem. Even with today's low bandwidth Internet video of 400 kbps, the *YouTub*e live concert needed more than 280Gbps server and network bandwidth. Akamai, the largest commercial *CDN* service provider reports a peak aggregate capacity of 200Gbps with its tens of thousands of servers [1]. Actually, the *Google* and *Akamai CDN* infrastructures are not necessary to handle live video stream effectively. Instead Peer-to-peer (*P2P*) video streaming systems can handle it effectively and far cheaper because peers are serving most of the video to others. P2P file downloading and streaming applications have recently attracted a large number of users on the Internet. The main difference between P2P file sharing and P2P streaming applications is the instant time when content is used. In P2P file sharing, content is completely transferred before files are opened. In contrast to P2P streaming where content is decoded, played immediately and later (probably) is discarded. Currently, several P2P video streaming systems have been deployed to reduce server cost. They are classified into two categories: *live* and *on-demand video* streaming systems. The live video streaming systems disseminate live video contents to all peers in real time. *Zattoo* was the largest Internet-based P2P live video streaming network in *Europe* [18]. In early 2006, the authors of [17] performed measurement study of a

P2P live video streaming system. In this study, 200,000 simultaneous users watched the live broadcast of 40hour event at bit rates from 400 to 800 kpbs. The aggregated required bandwidth reaches 100 gigabits/sec, while Akamai reportedly has roughly 300 gigabits/sec bandwidth at the end of year 2006. On the other hand, the on-demand video (*VoD*) streaming system enables peers to enjoy the flexibility of watching. It realizes the goal of watching whatever you want, whenever you want [34]. One of the most popular *VoD* platforms on the Internet is *YouTube* that hosts (mostly short) user generated content. There are also other *VoD* platforms focusing on full-length movies and TV programs, such as Hulu, BBC iPlayer, and Maxdome.

The current *P2P-VoD* systems cannot realize such goals and provide poor performance. For example, if a peer requests a video server to download one hour MPEG1 video (i.e. 540MB) using modem 28.8Kbps, 10Mbps then it has to wait 42 hours, 7.2 minutes to watch the video. Enormous television networks, news sites, and video sharing sites such as YouTube and Google Video provide *VoD* [14]. Most of the *VoD* being delivered today is short-length, low bit-rate clips. For example, YouTube videos today are typically less than 10 minutes in length and have a bit rate, under 200 kbps. In the near future, we expect a high demand for higher bit rate (potentially DVD quality) and longer videos (including full length movies) streamed over the Internet. *P2P-VoD* is a new challenge for the P2P technology. Moreover, *P2P-VoD* has less synchronous in the peers sharing video content compared to streaming live video content. Consequently it is more difficult to reduce load at server while maintaining the streaming performance. Thus, to reduce load at video server this paper presumes that each peer contributes a small storage for replicating video content. In addition, the video server divides the video data into segments. Each segment is divided into timeslots. Each timeslot maintains video data of few seconds video (eg. 25-30 sec).

This paper reviews the state-of-the-art of peer-to-peer technologies for video streaming, and presents taxonomy of various solutions that have been developed. The contribution of this paper is the design of P2P Video Overlay Network (*P2P-VON*) architecture. *P2P-VON* is a self-organized P2P overlay network. It organizes peers watching the same video into clusters. Each cluster contains peers with close playing points and overlapped buffer windows. Each peer identifies its neighbors based on and relative to their playing positions and its own playing position. Peers in the same cluster share a specific video segment. Moreover, this paper proposes an efficient algorithm to discover and share the video contents. It enables peer to exchange request messages with its neighbors and then schedule which video timeslots are to be fetched from which peer.

The remainder of this paper is organized as follows. Section two presents taxonomy of video streaming systems. Section three surveys the emerged approaches to construct video streaming P2P overlay networks.

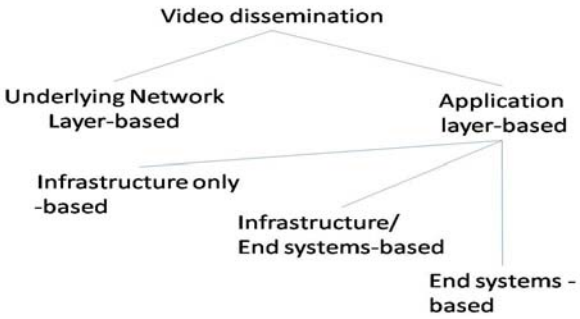

Figure 1. Video streaming taxonomy.

Section four discusses the proposed video overlay network, the step by step construction scheme, and the discover to exchange schedule algorithm. Simulation setup and results have been presented and discussed in section five. Finally this paper draws the conclusion.

II. TAXONOMY OF VIDEO STREAMING

There are two alternatives shown in Fig. 1 to disseminate video streams in the Internet which are underlying network layer-based and application layer based. The underlying layer-based video streaming has been implemented at the IP layer and are known as IP multicast model. It is the most efficient solution to the communication needs of multicast groups with a large number of members [9]. However, the deployment of IP multicast remains limited because of technical and non-technical barriers [10].First, it violates the stateless architectural principle, and introduces high complexity and severe scaling constraints at the underlying network layer [32]. Second, it requires substantial infrastructure modifications, and complex modifications to IP routers' software. The second alternative for disseminating video streams has been implemented in the application layer. It moved the multicast functionality away from routers towards end systems [13], [24], [6], [5], [25]. These approaches employed *unicast* IP service in multicasting. They implemented group membership, multicast routing, and packet duplication at end systems. Moving multicast functionality to end systems, instead of routers, engages performance penalties such as physical link stress [25]. The application layer-based includes three approaches *infrastructure only-based*, *infrastructure/end systems-based* and *end systems-based*. In infrastructure only-based, an organization uses *CDN* to deploy proxies at strategic locations on the Internet. These proxies broadcast packets to the nearby end systems using unicast. There are two approaches to build *CDNs*: overlay and network approaches [19]. The infrastructure only-based pursues the overlay approach. In the overlay approach, application-specific servers and caches at several places in the network handle the distribution of the video streams. Most of the commercial *CDN* providers such as *Akamai*, *AppStream*, and *Limelight* Networks follow the overlay approach for *CDN* organization. The infrastructure only-based approach is performance-effective but is also very expensive.

The infrastructure/end systems-based is an approach to video streaming that combines aspects of infrastructure only-based and peer-to-peer video streaming. Like infrastructure only-based video streaming, the infrastructure/end systems-based seeks to complement rather than replace the traditional client-server framework. Specifically, it considers the problem that arises when the server is overwhelmed by the volume of requests from its clients. For instance, a news site may be overwhelmed because of a large "*flash crowd*" caused by an event of widespread interest. Most of the famous systems that follow the infrastructure/end systems-based approach are *CoopNet* [23] and *Hybrid* video streaming scheme [30], *PROP* [31], etc. The infrastructure/end systems-based approach is cost-effective but not scalable due to limited storage and bandwidths of centralized broadcast servers that use unicast for the last hop to peers.

The end systems-based video streaming systems push the system's functionalities such as administration, maintenance and system operations (streaming) to end-systems (peers). They are referred as a peer-to-peer multicast, overlay multicast, or application-level multicast in the literature. They need to build a self-organized overlay networks that efficiently disseminate video streams. The P2P overlay networks are classified into three different approaches: *tree-first* (e.g. *ALMI* [25], *Yoid* [13], *mesh-first* (e.g. *Narada* [6], *Scattercast* [5]) and *implicit* (e.g. *CAN* [29], Community Overlay Network *CON* [27], [26] approaches. Unfortunately, although these overlay solutions alleviate some of the problems associated with IP multicast, they are often limited in their capabilities. For example, *Endsystem Multicast* constructs overlays entirely out of unicast connections between end-systems, and hence can scale to only small and medium sized groups. This paper proposes a novel overlay network that can scale to large sized groups.

A. *P2P video streaming issues*

This section discusses the P2P video streaming issues that must be addressed. The P2P video streaming systems require organizing peers into special and efficient overlay networks. These overlay networks should address the requirements to disseminate/share video streams to large scale (thousands of peers) while satisfying the real time and *QoS* constraints. Thus to construct and maintain an efficient overlay network, it is required to follow the following criteria:

1. The overlay network must be scalable to disseminate/share video to thousands of peers with reasonable overhead.
2. It should be efficient in disseminating video with high bandwidth and low latencies.
3. The overlay network should be resilience and robust to the dynamic changes such as join, leave, or fail.
4. The overlay network should be self-organized through applying distributed algorithms to build itself without help of central network administration.
5. Since the video streaming system relies on peers contributing heterogeneous bandwidth, it is required to achieve proper bandwidth sharing among incoming (and outgoing) connections to (from) individual peers. Thus, peers can receive different qualities of video based on their capabilities. To accommodate the bandwidth heterogeneity among peers a *Multiple Description Coding* (*MDC*) is used. In *MDC*, a stream is encoded into multiple sub-streams called *description*. Each description can be independently encoded.
6. The overlay network should satisfy the locality awareness requirement to improve the high level routing and information exchange in the application layer. This paper fulfills this requirement by organizing the overlay network into clusters.

It is clear that, the overlay construction and maintenance techniques are very critical and need to be investigated in details as follows. The following section assesses the emerged proposals for p2p video streaming overlay networks.

III. VIDEO STREAMING P2P OVERLAY NETWORKS

This section surveys the emerged approaches to construct overlay networks. In particular, the approaches can be classified into three categories, namely mesh-based, tree-based and data-driven.

A. *Mesh-based approach*

In this approach, peers form a randomly connected overlay, or a *mesh*. Each peer maintains a set of neighbor nodes. Certain number of neighbor nodes are categorized as parents (*i.e.*, incoming degree) and others serve child peers (*i.e.*, outgoing degree). When a peer arrives it contacts a bootstrapping node to receive a set of peers that can potentially serve as parents. The bootstrapping node maintains the outgoing degree of all participating peers. Then, it selects a random subset of peers that can provide slots for new child peers in response to an incoming request for parents. The mesh-based approach facilitates content delivery in either bidirectional or unidirectional fashion. It employs a content delivery that effectively utilizes the outgoing bandwidth of participating peers as the group size grows. This content delivery couples push content reporting with pull content requesting. Individual peers periodically report their newly available packets to their child peers and request specific packets from individual parent peers. For utilizing the available bandwidth effectively, it is required to implement a *packet scheduling algorithm* that determines the number of requested packets from parents.

Several mesh-pull P2P streaming systems have been successfully deployed to date, accommodating tens of thousands of simultaneous users. The pioneer in the field, *CoolStreaming*, reported that more than 4,000 simultaneous users in 2003 [2]. More recently, a number

of second-generation mesh-pull P2P systems have reported phenomenal success on their Web sites, advertising tens of thousands of simultaneous users who watch channels at rates between 300 kbps to 1 Mbps [17]. These systems include *PPLive*, *PPStream*, *UUSee*, *SopCast*, *TVAnts*, *VVSky*, and *PRIME*, [21].

In P2P live streaming, peers are interested in the similar part of the video content. While due to the asynchronous nature of *P2P-VoD* service, the peers are interested in different parts of video content at any given moment. In *P2P-VoD*, the video is divided into segments that regularly exchanged between peers according to a certain download strategy [34]. For example, *PONDER* is performance aware *P2P-VoD* system that incorporates the mesh-based P2P downloading into the traditional server-client video-on-demand service [33].

B. Tree-based approach

In this approach, an overlay construction mechanism organizes participating peers into multiple trees. Each peer determines a proper number of trees to join based on its access link bandwidth. To minimize the effect of churn and effectively utilize available resources in the system, participating peers are organized into multiple *diverse* trees. Nodes in the tree have well defined relationships, for example, "*Parent-child*" relationships in trees. Toward this end, each peer is placed as an *internal* node in only one tree, and as an *external* (or leaf) node in other trees. Then, each description of *MDC* encoded content is delivered through a specific tree. The content delivery employs push mechanism where internal nodes in each tree simply forward any received packets for the corresponding description to all of their child nodes. Further, the tree structure must be maintained, as nodes leave and join the group. The failure of nodes that are higher in the tree may interrupt delivery of data to a large number of peers, and result in poor transient performance. The authors in [3], [23] proposed central construction tree algorithms that constructed multiple balanced, stable and short trees. Finally, when constructing tree-based structures, loop avoidance is an important issue that must be addressed. Therefore, the main component of the tree-based P2P streaming approach is the tree construction algorithm. This approach has a number of drawbacks as follows. First, it does not utilize the outgoing bandwidth of a large fraction of nodes (i.e. leaves). Second, the received bandwidth is limited by the minimum bandwidth on the path to the source and any loss in the upper level of the tree reduces bandwidth available to nodes lower in the tree. Finally, it has poor resilience to churn (a node departure results in the data stream being lost at all its descendants until the tree is fixed). Multiple tree-based approaches [3] aim at solving the first two problems, but do not solve the third one as they require more tree structures to be maintained. Motivated by these drawbacks, recent research has focused on constructing and maintaining the overlay topology in mesh-based. Richard John Lobb, et. al., [20] proposed an algorithm that optimized the topology to better exploit large bandwidth peers, so that they are automatically moved close to the source. This improves the chunk delivery delay so that all peers benefit, not just the high bandwidth ones. A key property of the proposed scheme is its ability to indirectly estimate the upload bandwidth of peers without explicitly knowing or measuring it.

P2Cast [16] was designed for distributing video content among asynchronous users, where every user act as a server while retrieving media content. The entire video is then streamed from the source server using this base tree; just the same way as in tree-based P2P live streaming. Peers joining the session join the base tree and retrieve the base stream from it. In contrast, new peers who missed the initial part of the video must obtain a patch directly from the server or other peers who have already cached the required content. Peers have two main functions: (i) *base stream forwarding*, where users participating in the tree-based overlay should be capable of forwarding the received base stream to others; and (ii) *patch stream serving*, where users cache the initial part of the video and serve the patch to latecomers.

C. Data-driven approach

In contrast to the tree-based approach, the overlay network designs in the ***data-driven*** approach do not construct and maintain an explicit structure for delivering video streams [2]. The streaming overlay is designed in the data-driven approach, where a node always forwards data to others that are expecting the data, with no prescribed roles like parent/child, and down-streaming /up-streaming, etc. In other words, it uses the availability of the video segments to guide the segments flow. Moreover, it needs a scheduling algorithm that struggles to schedule the segments that must be downloaded from various peers to meet the playback deadlines.

The data-driven approach employs gossip algorithms [11] as sample approaches to disseminate video segments without explicitly maintaining a structure. In gossip algorithm, a peer sends a newly generated message to a set of randomly selected peers; these peers in turn send same message to other peers until the message is spread to all peers. Clearly, gossip is not appropriate for video streaming because its random push may cause significant redundancy with the high-bandwidth video. *DONet* [2] adopts pull-based technique to avoid redundancy, as peer pulls segments from other peers only if it does not already hold it. *Cool-Streaming* overlay is robust to failures as any segment may be available at multiple partners [2]. Gossip protocols incur an extra factor of *log n* overhead as compared to tree based overlay networks because the number of transmitted messages is *O(n log n)* in the former as compared to *O(n)* in the latter. Storage-assisted Data-driven overlay Network (*SDNet*) enhanced the *DONet* [4]. Its routine video distribution is based on the buffer overlapping mechanism and gossip protocol that support the efficient *VoD* operations.

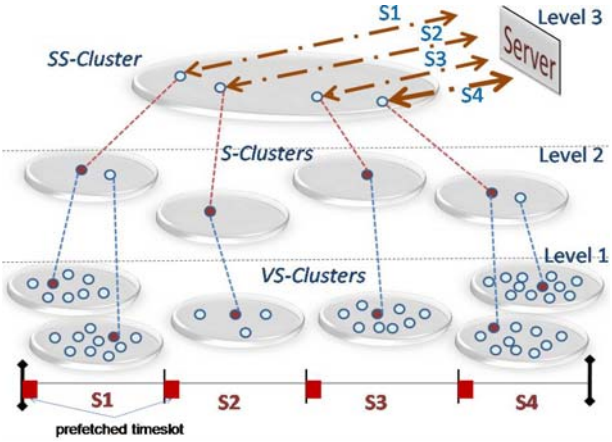

Figure 2. P2P Video overlay network architecture.

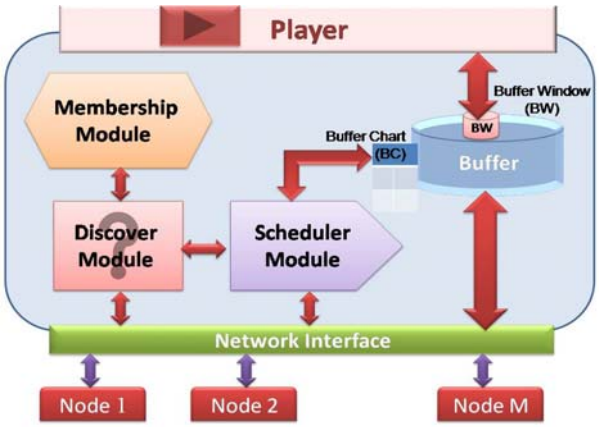

Figure 3. Node architecture in the P2P-VON.

IV. P2P-VON ARCHITECTURE

This section describes the proposed P2P Video Overlay Network (*P2P-VON*) architecture, its node architecture and construction algorithm. In this paper, video server contains some modules that enable it to split and decode video contents as follows. First, Video splitter that works with the help of the content decoder to convert the video contents into series of *timeslots* (i.e. atomic playable parts of the video). The *timeslot* encapsulates the video data contents of *25-30*sec. Second, Video content decoder is a part of the *codec* that knows how to split the video file into independent playable series of timeslots. In *P2P-VON*, each node retains a *buffer window* (*BW*) that contains at most *BW* timeslots. The Windows media player needs to buffer one minute of video contents before playing. Consequently, each node in *P2P-VON* maintains a buffer that can store a segment of video data. For example, the segment encapsulates the *MPEG1* video contents of five minutes (i.e. 45MB) at most. The availability of the video timeslots in the buffer of a P2P-*VON* node can be represented by a *buffer chart* (*BC*) that represents the availability of the timeslots in its buffer. Periodically, each *VON* node exchanges request messages with its neighbors and then schedules which video timeslots are to be fetched from which node as shown in the *discover to exchange schedule algorithm* in the following sections. Thus in *P2P-VON* each node can discover the other nodes that are maintaining the next minute video contents, download and buffer it during the playing of the preceding minute. The proposed overlay network organizes the *P2P-VON* nodes into Video Segments Clusters (VS-cluster) based on the clustering parameters that are defined as follows.

Definition *1* (***<α, β, μ>-Clustering***): *Given a graph G = (V, E, <α, β, μ>), where vertex* $v \in V$ *, edge* $e \in E$ *. The 3-tuble <α,β, μ> is an integer and called the clustering parameters.*

It clusters the graph G into connected sub-graphs called *Video Segment Clusters* (*VS-Cluster*). The physical distance between any two end-nodes in each *VS-Cluster* is less than or equal β physical hops. The distance between playing points of any two vertices in the *VS-Cluster* is less than α (*segmentation parameter*). The number of logical hops between any vertices in *VS-Cluster* is bounded by μ.

Definition 2 ***(VS-Cluster)***: *The VS-cluster is defined by a subset of nodes which are mutually playing same video and the differences between their playing points are less than α.*

The 3-tiers architecture of *VON* is shown in fig. 2 and will be explained as follows. It is composed of multiple VS-Clusters VS_k ; *k=1, …,ω* in level 1. Where $|VS_j|$ is the size of a VS-Cluster *j*. Each VS_k has a *supervisor* node called s_k. For pair of peers $i, j \in VS_k$ the distance between their current playing point positions is $|Pos_i - Pos_j| \le \alpha$ where Pos_i and Pos_j is the playing position of peer *i* and peer *j* respectively as shown in fig. 2. Moreover, all peers belongs to the same *VS-Cluster* should store the first timeslot so called a *prefetched timeslot* as shown in fig. 2. The *prefetched timeslot* enables the user seeks (jumps) to a new playing point at different segment (i.e. another *VS-Cluster*) to be satisfied immediately. Consequently, the *prefetched-timeslot* reduces the time from when the user initiates a seek to when the video begins playing at the new offset. The *distance* between two vertices *i* and *j*, *d(i, j)*, is the number of logical hops of a shortest path between *i* and *j*. Its maximum value over all pair of vertices, $D_\tau^{(k)} = max\ \{d(i, j);\ \forall\ i, j \in VS_k\}$, is the *diameter* of the *VS-Cluster* VS_k at instance of time *τ*. Where $D_\tau^{(k)}$ is less than or equal the upper bound diameter *μ*. In case of the number of peers watching the same segment is large, the *VS-cluster* will be organized into multiple of *VS-clusters* based on β, and μ. For example, fig. 2 shows two *VS-Clusters* associated with each video segment S1 and S4. Level 2 contains Supervisor-Clusters that are called *S-Clusters*. Each *S-Cluster* contains the supervisor nodes of the *VS-Clusters* that shared the same video segment. The supervisor nodes elect a node to be a *super-supervisor*. The number of the *VS-Clusters* in level one is denoted by ω. In level 3 a cluster that contains all

super-supervisor nodes is constructed and called Super-Supervisors-Cluster *SS-Cluster* as shown in fig. 2.

A. P2P-VON: Node Architecture

Each node has a unique identifier. Such as its IP address and maintains a membership list containing the list of identifiers of its neighbors. We adopt pull-based technique to avoid redundancy, as node pulls video timeslots from other nodes only if it does not already hold it. For that reason, each node maintains a set of nodes (partners), and periodically exchange video timeslots availability information with the partners. Fig. 3 demonstrates the *P2P-VON* node architecture. It has three main modules: (1) *Membership module*, which helps the node to maintain a list of neighbors in its *VS-Cluster*. Moreover, it helps when the user seeks to a new playing point located at different segment. Consequently, it is responsible for node migration from current *VS-Cluster* to another one; (2) *Discover module*, which enables a node to discover other nodes that cache the required video timeslots to play. It has dual functions send and receive video timeslots requests; (3) *Scheduler module*, which schedules the transmission of video timeslots. Each node can be either receiver, supplier, or both depending dynamically on this video timeslots availability.

B. *P2P-VON: Step by step Construction Scheme*

This section broaches the step by step self-construction algorithm for *P2P-VON*. A video broadcasting system typically has a single dedicated video source, which is assumed not to fail. The address of the source is known in advance, serving as a rendezvous for new end-users to join the video streaming session. The video server maintains the list of super-supervisor peers, which are located in *SS-Cluster*. When an end-user node wants to watch certain video, it contacts the video server with the position ρ it wishes to start receiving. It calls *Join_VON(ρ)* process that performs the following steps.

i. If new peer *X* wants to watch the video at certain position belongs to a segment and no other peers are watching it at the moment, the video server replies with " no super-supervisor peer". Then peer *X* calls *join(SS-Cluster)* and constructs a new *S-Cluster* in level 2 as shown in fig. 4. It also constructs a new *VS-Cluster*, such as VS_7 as shown in fig. 4. Finally, the video server starts to broadcast the contents of the video segment S_5 to node *X*. The number of supervisor peers in *S-Clusters*, ω increases by one. For example, in fig. *4* the node *X* becomes a supervisor peer of the new *VS-Cluster* VS_7. This new *VS-Cluster* is denoted by $VS_{\omega+1.}$

ii. Otherwise new node (e.g. *Y)* wants to watch the video at certain position ρ belongs to a segment and there is some peers watching the same segment, then the video server finds a super-supervisor peer *say, H* of the *S-Cluster* where ρ position held in its video segment (e.g. S_4) as shown in fig. 4.

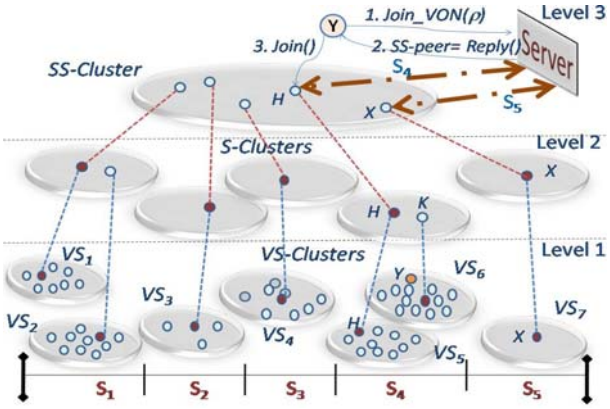

Figure 4. Video overlay network step-step construction.

a. The peer *H* is also the supervisor of the *VS-Cluster* VS_5. It checks if $[((D_\tau^{(j)}+1)\le \mu)$ and $(d(Y, t_j) \le r; \delta=2r)]$ then node *H* calls *Join_VSCluster*(VS_j) process to join the *VS-Cluster* VS_j (e.g. $j=5$). The *Join_VSCluster*(VS_j) process chooses χ random nodes $m_l \in VS_j; (l = 1,..., \chi)$ and then inserts node *X* between each node l_i and its successor node $(m_l \rightarrow n_s^{(i)})$ in the *i-th* Hamilton Cycle *(HC)* similar to our previous work in [28].

b. Otherwise, if $[d(Y, t_j) > r$ *or* $((D_\tau^{(j)}+1)> \mu)]$, the supervisor node *H* broadcasts a join request message in the *S-Cluster* and then waits *time-out* σ period for a reply from any supervisor peer in the *S-Cluster*. There are two cases:

1. No response is received within the time-out interval σ, then node *Y* creates its own new *VS-Cluster* and becomes a supervisor peer. The number of supervisor peer (*S-Clusters*) ω increases by one. For example, node *Y* becomes a supervisor peer of the new *VS-Cluster* VS_8.
2. If *H* receives multiple of replies then it selects the t_s with the smallest distance to *X* where $d(Y, t_s) \le r$, $((D_\tau^{(s)}+1)\le \mu)$ and the minimum $|VS_s|_\tau$. Then node *Y* calls *Join_VSCluster*(VS_s) process to join the *VS-Cluster* VS_s. . In addition, the new joined node should store the first (*prefetched) timeslot* from its neighbors. For example, node *Y* joins VS_6 and stores the *prefetched timeslot* of S_4 as in fig. 4.

C. Discover-to-Exchange Schedule Algorithm

In the P2P-VON, each node needs to carry out the following discover-to-exchange (D2E) schedule algorithm to discover the other nodes that buffered the following video minute to play.

1. Periodically, the scheduler module at any node in the *P2P-VON* (e.g. *X*) finds out what the next video timeslots to play and emits signal to the *discover module* to discover the others nodes in its *VS-Cluster* that hold the required video timeslots.
2. The *discover module* at *X* creates a *request message* (*Rm*) as shown in fig. 5. It contains IP of node X and

Figure 5. Request message (Rm) format.

movie identifier (*MID*), index of the current timeslot (*ICS*) and *index of the requested timeslot* (*IRS*).

3. As soon as, the *discover module* acquires the neighbors list of node *X* from the *membership module*, it sends out the *Rm* to the neighbors of the node *X*.
4. The *discover module* at each node (e.g. *Y*) received the *Rm*. To avoid duplication, each node monitors the received *Rms* and then forwards only one copy of the duplicated ones as follows:
   a. If node *Y* does not receive such *Rm* before it forwards it to its schedule module.
   b. Otherwise, Y omits the received *Rm*.
5. Then, the *schedule module* (see psuedo code) at node *Y* checks if the index of the required video timeslots with certain movie *MID* is located in its buffer chart (*BC*) or not. There are two cases as follows:
   a. If the required video timeslots is maintained at node *Y*, then the *schedule module* coordinates with the discover module to forward its IP to the requester node (*X*).
   b. Otherwise, the *discover module* at *Y* obtains its neighbor list and then forwards the request message to its neighbor nodes within its *VS-Cluster*.
6. The *discover module* at the requester node *X* waits for a *time-out* to receive reply message(s) containing the IP address of the node(s) that maintain the required video timeslot (*IRS*). In fact there are two cases as follows:
   a. Node *X* collects the received replies and selects the most closest node to download the required video minute. Then, it downloads the required video minute.
   b. Otherwise, *X* does not receive any reply message within the time-out, then it forwards the request message to the supervisor node of its *VS-Cluster*. Accordingly, the supervisor node forwards the request message to the super-supervisor node in *SS-Cluster*. Then, the super-supervisor node downloads the required video minute from the video server and forwards it to the supervisor node that consequently forwards it to the node *X*.

It is clear that any node in the *D2E* schedule algorithm may send request messages within the *VS-Cluster* with size $|VS_j|$ to at most μ hops. Similarly in unstructured P2P network [34] queries on a tend to have lookup complexity of the order of μ, or O(μ), hops.

```
Discover_module()
BEGINE
 CASE status OF
 Requester:
       //Node who init a request for a timeslot IRS
   MSG = CALL create_MSG(MID,ICS,IRS)
     N =  CALL Requester.neighbors()
        // N is the set of neighbors of requester node
    CALL Send(MSG,N)
         // Send request MSG to set of neighbors
   R= CALL Wait_reply(t);
        // Where  t : is the timeout wait for replies
        // and
        //         R : is the set of relies messages.
   IF (R ≠ ϕ ) THEN
    CALL Requester.download(IRS,
                                Closest(R))
   ELSE // No reply received
   Supervisor=
Requester.MyVS_Cluster_supervisior()
    CALL Send(MSG, Supervisior)
   ENDIF
  Replier: // Node who received request message
           // for timeslot IRS
    RMSG= CALL receive()
          // RMSG is the received message
    Located = CALL schedule(RMSG)
          // schedule module return true if
          // RMSG.IRS ∈ Replyer.BC(buffer chart)
    IF (Located)THEN
          Reply_with_ip(Replier.ip)
    ELSE
          M= CALL Replyer.neighbors()
           //M:set of neighbors of replier node
          CALL Send(RMSG, M)
      // forward the received message to its neighbors
   ENDIF
    ENDCASE
END
```

IV. PERFORMANCE EVALUATION

This section evaluates the performance of the proposed solution through simulation experiments. In order to evaluate the performance, this paper uses the following performance metrics.

- *Server stress* measures the average amount of bandwidth used at the server.
- *Initiate time* is the time that peers have to wait from first access to the video until receiving the video and playing it back.
- *Seek time* is the time from when the user initiates a seek to when the video begins playing at the new offset.
- *Average Network workload* is the sum of all link workload $W = \sum_i w_i$, where the link workload of link *i*, w_i is the product of link utilization and its capacity.
- *Timeslot cost* is defined as the total number of timeslots fetched from the source server. It is

represented by the average load generated on the server for each played timeslot.

- Segmentation fraction γ is the ratio of segment size to the total video length ϖ. The impact of γ should be studied. (e.g. suppose one hour video, segmentation fraction is 20% of the total video then video is divided into 5 segments and each segment is 12min). As γ increases, the *segmentation parameter* $(\alpha = 1/\gamma)$ decreases. Consequently, the size of each *VS-Cluster* increases and more peers share the same segment.

The simulator operates in discrete intervals of time (called timeslots), each of which is 25s in duration. Assume the streaming rate is 500kbps and hence the amount to be transferred in a timeslot is 25s * 500kbps $\approx$ 1.5MB. Moreover, the simulation supposes a video of ϖ = 60min in duration and hence, the entire file consists *of 60min / 25s = 144* timeslots, which we have rounded to 150 timeslots in this experim*ent. For e*xample, when $\gamma = 20\%$ then the video of 60min is divided into 5 segments, each segment is 12min and around 30 timeslots. Each peer in the *VS-Cluster* should store the first timeslot of each segment.

To evaluate performance, we developed a simulation a 3-level network topology. The top two levels are generated by *GT-ITM* [12] with 100 nodes connected by core links. The simulation used to provide only one video and consequently, sets the capacity of each core link to be 15Mbps that can support up to 15 simultaneous streams. The link delay was randomly set between 1 and 10ms for each edge link. The capacity of each edge link is 5Mpbs in this simulation. The network consists of one transit network including 6 nodes and 10 stub domains. We employed the shortest path algorithm to assure efficient routing. This paper assumes that each node is an abstraction of a local network that can host an unlimited number of peers. Each peer is equally likely to be placed at any node. More than one peer could be reside at the same node. The average arrival rate of the peers is λ. Each client offers 18MB physical data storage capability. Set video's streaming rate to 500kbps then each peer is able to store 5 minutes' video data. In this paper, the simulator assumes that the video playback rate is constant bit rate. Also, it assigns a bandwidth to each link in terms of the number of playback rate a link can support.

A. Simulation Results

This section discusses the simulation results as follows. Fig. 6-a shows timeslots cost χ versus the simultaneous viewers σ in the *VS-Cluster*. It compares the proposed *P2P-VON* and client-server model. In the client-server model, we assume that every peer only obtained timeslots from source server. Consequently, its timeslot cost is 1.0. However, in *P2P-VON* each peer caches the recently watched timeslots in its buffer. The simulation results of the timeslots costs are shown in fig. 6-a. As the number of viewers exceeds 1, the timeslots cost is improved. This shows that the proposed *P2P-VON* can significantly offload source server as the number of the simultaneous viewers increases. When the number of simultaneous viewers reaches 6, the timeslots cost of the *P2P-VON* is about 0.19. That means there are only 19% of timeslots fetched from source server. There are two reasons behind fetching timeslot from the source server. First, there is no peer belongs to the *VS-Cluster* caches the scheduled timeslot. Second, peers in the neighborhood cache the scheduled timeslot, but there is no available bandwidth to any of them. It is clear from fig. 6-a that timeslot cost induced by the insufficient bandwidth is decreasing as the number of simultaneous viewers is increasing. Fig. 6-b illustrates the average number of nodes needs to be connected within the *VS-Cluster* for different γ (*segmentation fraction*) in *P2P-VON*. As γ increases, the connectivity within the *VS-Cluster* increases (i.e. the average number of nodes needs to be connected within the *VS-Cluster* increases). Obviously a large γ in *P2P-VON* allows more users to be served, however it also leads to a large joining delay. The required storage placed on each node also increases. Consequently, it is significant to carefully choose the *segmentation fraction* such that most benefits can be obtained without overwhelming the node. Fig. 6-c illustrates the server stress vs. the peers arrival rate using different techniques. The *unicast-server* technique provides a stream for each unique user. It is most stressed technique. However, the proposed *P2P-VON* technique lessens the workload placed at the server. Fig. 6-d depicts the average network workload vs. the peers arrival rate with different values of the segmentation fraction. It is clear that as peer arrival rate increases, more peers joins the same *VS-Cluster* and thus increases the average network workload. Moreover, fig. 6-d shows the differences in the average workload on the network is insignificant as a function of the segmentation fractions. In addition, fig. 6-e shows the changes of average latency with the increasing of the segmentation fraction. More neighbors will definitely help, as the needed timeslots can be downloaded in parallel. Clearly, the initiate time and setup time drop with the increasing of the segmentation fraction. Finally, fig, 6 (a-e) indicates that the proposed *P2P-VON* alleviates the server stress and does not significantly load the network.

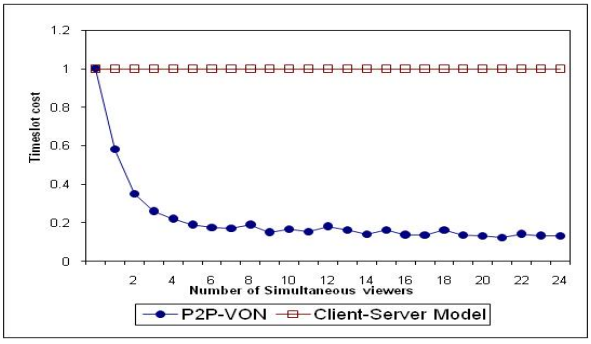

Figure 6-a. Timeslots cost χ versus the simultaneous viewers σ in the VS-Cluster.

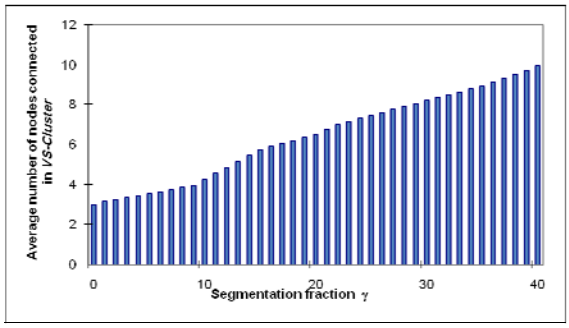

Figure 6-b. Average number of nodes needs to be connected within the VS-Cluster for different γ (segmentation fraction) in P2P-VON.

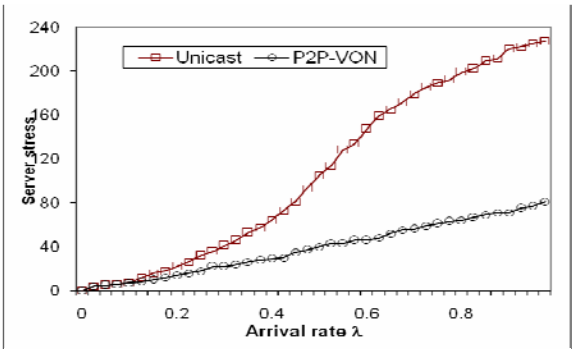

Figure 6-c. Server stress vs. peers arrival rate for different techniques.

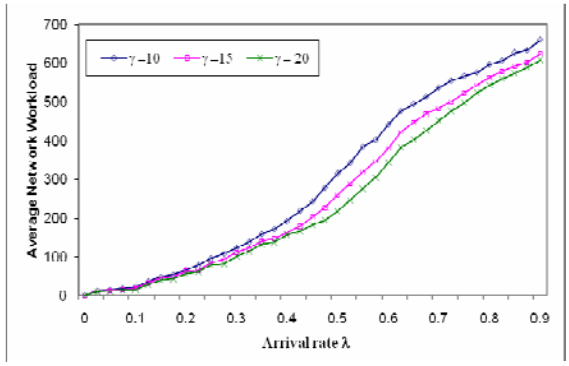

Figure 6-d. Average workload on the network as a function of the segmentation fractions.

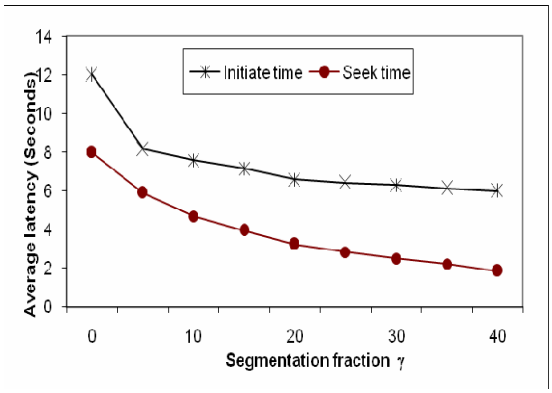

Figure 6-e. Alleviation of the server stress and the load of the network by the proposed *P2P-VON*.

VI. CONCLUSION

This paper presents an overview of the video streaming applications and introduces taxonomy of video streaming systems. Moreover, it presents a novel overlay network construction technique. This technique organizes the overlay network into 3-tier architecture of clusters. It clusters the *VON* into *VS-Clusters* based on three parameters (*number of physical links, logical diameter, and sharing video segment*). Implementing and maintaining these parameters improves the performance of video on-demand and video dissemination. The simulation results show that proposed *P2P-VON* scales better than conventional client-server *unicast* for the *VoD* service. Moreover, it demonstrates the proposed *P2P-VON* alleviates the server load with insignificant network load.

REFERENCES

[1] Akamai, *Akamai Technologies Inc.* Retrieved from www.akamai.com, 2007

[2] Zhang, X., et. al., "*DONet/CoolStreaming: A data-driven overlay network for live media streaming*", In *Proc. of INFOCOM'05*. Miami, FL. 2005.

[3] Castro, M., Et. al., " *SplitStream: High-bandwidth multicast in cooperative environments*", In *Proc. of ACM SOSP'03*. New York, 2003.

[4] Changqiao X., "*Distributed Storage-Assisted Data-Driven Overlay Network for P2P VoD Services*", IEEE Trans. On Broadcasting, Vol 55. No. 1, 2009.

[5] Chawathe, Y., "*Scattercast: An Architecture for Internet Broadcast Distribution as an Infrastructure Service*", (Ph.D Thesis). Univ. of California, Berkeley, 2000.

[6] Chu, Y. H., Et al., "*A case for end system multicast*", In Proceedings of ACM Sigmetrics, 2000.

[7] Cisco, "*Visual Networking Index: Forecast and Methodology", 2010-2015*. White Paper, 2011. URL http://www.cisco.com/en/US/ solutions/collateral/ns341 /ns525/ns537/ns705/ns827/white_paper_c11-481360_ ns827_Networking_Solutions_White_Paper.html.

[8] ComScore, "*ComScore measuring the digital world*", Retrieved from http://www.comscore.com, 2009

[9] Deering, D., "*Multicast Routing in Internetworks and Extended LANs*", Proc. of the ACM SIGCOMM, 1998.

[10] Deering, S., "*Multicasting routing in a datagram internetwork*", (Ph.D. Thesis), Stanford University, 1991.

[11] Eugster, P., et. al., "*From epidemics to distributed computing*", IEEE Computer. *Accustream iMedia research*, 2009, http://www.accustreamsearch.com

[12] E. Zegura et al. ,"*How to Model an Internetwork*", Proc. Joint Conf. IEEE Computer and Comm. Societies (INFOCOM), IEEE CS Press, 1996, pp. 594–602, 1996.

[13] Francis, P. ,"*YOID: Extending the Multicast Architecture*" White paper. http://www.aciri.org/yoid (1999).

[14] Google video. (n.d.). *Google video.* Retrieved from http://http://video.google.com/

[15] G. Marfia, et al., "*P2P Streaming Systems: A Survey and Experiments*", 3rd STMicroelectronics STreaming Day (STreaming Day'07), Genoa, Italy, September 14, 2007

[16] Guo, Y., et. al. , "*P2Cast: Peer-to-peer patching scheme for VoD service*", Proceedings of the 12th International Conference on World Wide Web, 301-309, (2003).

[17] Hei, X. et. al. , "*A Measurement Study of a Large-Scale P2P IPTV Systems*", IEEE Trans. On Multimedia, 2007.

[18] Hyunseok C., et. al. , "*Live Streaming Performance of the Zattoo Network*", In the 9th ACM IMC Conf., 2009.

[19] Lazar, I., & Terrill, W., "*Exploring Content Delivery Networking. IT Professional*", 3(4), 47-49, 2011.

[20] Lobb, R. J., et. al., "*Adaptive overlay topology for mesh-based P2P-TV systems*", Retrieved in 2009 from

http://www.informatik.uni-trier.de/~ley/db/conf/nossdav/nossdav2009.html#LobbSLMM09

[21] Magharei, N., et.al., "*Prime: Peer-to-peer receiver-driven mesh-based streaming*", Proc. of IEEE INFOCOM, 2007.

[22] Mukaddim P. and Rajkumar B. , "*A Taxonomy of CDNs", Content Delivery Networks*", Springer, Germany, 2008.

[23] Padmanabhan, V. N., et. al., "*Distributing Streaming Media Content Using Cooperative Networking"*, Miami Beach, FL: ACM NOSSDAV, 2002.

[24] Zhuang, S. Q., et. al., "*Bayeux: An architecture for scalable and fault-tolerant wide-area data dissemination*", *Proc. of NOSSDAV'01*. New York, 2001.

[25] Pendarakis, D., "*ALMI: an application level multicast infrastructure*". Proc. of the 3rd Symp. on USITS, 2001.

[26] Ragab, K., "*Autonomous Decentralized Community Communication Technology for Assuring Information Dissemination*", (Ph.D. Thesis). TITCH, Tokyo, 2004.

[27] Ragab, K., & Mori, K., "*ACIS-Hierarchy: Enhancing Community Communication Delay for Large-Scale Information Systems*". IEICE Trans. on Comm., E87-B, N(7), 2004.

[28] Ragab, K., et. al., "*Autonomous Decentralized Community Communication for Information Dissemination*", IEEE Internet Computing Magazine, 8(3), 29-36, 2004.

[29] Ratnasamy, S., et al.,"*A Scalable Content- Addressable Network"*, In *Proceedings of SIGCOMM'01*., USA, 2001.

[30] Shan, Y., & Kalyanaraman, S., "*Hybrid Video "Downloading/Streaming Over Peer-to-Peer Networks*", Proc. of the Int. Conf. on Multimedia and Expo, 2003.

[31] Songqing, L. G., et. al., "*PROP: a Scalable and Reliable P2P Assisted Proxy Streaming System*", Proc. of the Int. Conf. on Distributed Computing Systems, 2004.

[32] Zhang, B., & Hussein, T. M., "*Forwarding state scalability for multicast provisioning in IP networks"*, In IEEE Communications Magazine, 46–51, 2003.

[33] Yang Guo, et al., "*PONDER: Performance Aware P2P Video-on-Demand Service*", IEEE GLOBECOM 2007.

[34] Yong Liu, et. al., "*A Survey on Peer-to-Peer Video Streaming Systems*", Peer-to-Peer Networking and Applications, 1(1):18–28, 2008.

Fuzzy Based Reactive Controller for a Small Mobile Robot Platform

RUSU Călin[1], BARA Alexandru[2], SZÖKE Enikö[1], DALE Sanda[2]

[1]Technical University of Cluj-Napoca, [2]University of Oradea
calin.rusu@edr.utcluj.ro, abara@uoradea.ro

Abstract — A complete mobile robot navigation system should integrate the local and global navigation systems. In the last decade many papers have been reported concerning the robotic navigation. Such a system can be separated into two levels of control: global path planning and local motion control. Local motion usually use sensory information to determine a motion that will avoid collision with unknown obstacles or obstacles whose position in the environment had changed. The paper deals with design procedures of a fuzzy based local obstacle avoidance system for an autonomous mobile robot using IR detection sensors. A differential small mobile robot is used to interact with an unknown indoor environment using a reactive strategy determined by sensory information. Fuzzy behavior was preferred for local navigation of a robot because this is well suited for uncertainty of the environment. A rule-based fuzzy controller with reactive behavior was implemented and tested for a mobile robot equipped with infrared sensors to perform collision-free navigation. The simulations and experimental results are developed and shown that the proposed architecture is not expensive and also provides an efficient and flexible solution for small indoor mobile robots.

Index Terms — Fuzzy controller, IR sensors, Differential mobile robot, Obstacle avoidance

1. INTRODUCTION

The mobile robots are used extensively in many fields, such as surveillance and/or intervention in hostile environments, or in the warehouses as a solution for inspection and operation. Autonomous navigation is related to the ability of a mobile robot to move in an unknown environment and to achieve a goal without hitting any obstacles. In the last decade a lot o researches covering aspects including, but not limited to *wall following*, *collision avoidance*, *corridor following* or *goal seeking* have been reported. The navigation system is essential and can be separated into two levels of control: *global path planning* and *local motion control*. Present paper deals with *local navigation system*. A reactive behavior navigation problem based on stimulus–response mechanisms was adopted. The mobile robot has the ability to navigate autonomously avoiding any type of obstacles. This reactive strategy is completely based on sensory information [1]. An absolute localization is not requisite and the structural modeling of the environment is unnecessary. The robot is expected to carry out only simple tasks through its sensory inputs as a set of stimulus–response mechanisms [9].

The control problem for the two-wheel mobile robots is how to independently control the left-wheeled motor and right-wheeled motor. Classical control strategy has proposed a simple PI feedback controller with feed-forward compensator for each of motor drive. This approach control method use extensively inverse kinematics and dynamics of the robot [2], [3]. In our application, we considered a small two wheels mobile robot. For this robot the use of mathematical model in order to control motion is too "expensive" in terms of computer processing power and memory [16].

In our paper the robot is guided by online sensor information acquired while navigation is performed. A behavior-based or sensor-based method have been adopted and implemented. This approach drives the robot by direct mapping between sensors and motors without building predefined environmental maps [6]. Fuzzy logic was preferred since unlike classical logic this is tolerant to imprecision, uncertainty and partial truth. Also the fuzzy controller is easier to be implemented for nonlinear models rather than other conventional control techniques. More than it, the fuzzy controller offers the possibility to mimic expert human knowledge [7].

The paper is organized as follows: Section 1 Introduction. Section 2: Mobile robot system presentation. Section 3: Rule based fuzzy navigation system. Section 4: Simulation and experimental results and Conclusions.

2. MOBILE ROBOT SYSTEM

For our experiments we have take full autonomous differential two-wheels mobile robot used for indoor environment. The robot has two identical parallel, wheels (attached to both sides of the platform) which are controlled by two independent DC gear motors. Also it assumed that each wheel is perpendicular to the ground and the contact between the wheels and the ground is pure rolling and non-slipping. The velocity of the center of mass of the robot is orthogonal to the wheels axis - *L*.

The center of mass of the mobile robot (*x, y*) is located in the middle of the axis connecting the wheels (*L*), likes in Fig. 1.

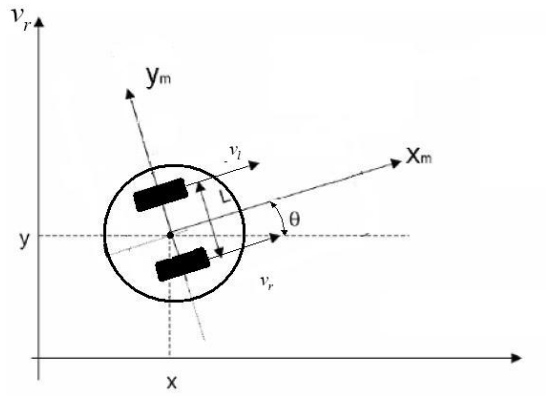

Figure 1. Kinematics model of the robot.

This robot is relatively easy to model and build. The linear left (right) wheel velocity v_L (v_R) is directly proportional to the angular velocity of the left (right) wheel (1), where *r* is the wheel radius.

$$v_L = r \cdot \omega_L$$
$$v_R = r \cdot \omega_R \quad (1)$$

The angular velocities (ω_l and ω_r) of the two wheels are independently controlled. This mobile robot is an under actuated system, because it has two inputs (*translational velocity* – v_c and *angular velocity* – ω_c) and three outputs (*center positions x, y and heading angle θ of the mobile robot on two dimensional Cartesian workspace*) [15].

$$v_c(t) = \frac{v_L + v_R}{2}, \quad (2)$$
$$\omega_c = \frac{v_L - v_R}{L}$$

where parameter *L* is a robot chassis radius. The kinematics for this robot is given by relation following relation.

$$\begin{bmatrix} v_x(t) \\ v_x(t) \\ \dot{\theta}(t) \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\cos\theta & \frac{1}{2}\cos\theta \\ \frac{1}{2}\sin\theta & \frac{1}{2}\sin\theta \\ \frac{-1}{L} & \frac{1}{L} \end{bmatrix} \cdot \begin{bmatrix} v_r \\ v_l \end{bmatrix}. \quad (3)$$

This relation is simplest and the most suitable for a small-sized battery-driven autonomous robot, and was used for simulation under Matlab-Simulink.

For allowing obstacle avoidance, the robot is equipped with a set of proximity infrared (IR) sensors - Sharp GP2D120. The problem of obstacle avoidance is a basic and crucial part of any autonomous navigation system. The sensors range finders are used to help in determining the correct direction for the robot to turn after it has detected a frontal obstacle. Each sensor is used to detect the obstacles by measuring the reflected light as is shown in Fig. 2. A light is transmitted from the IR emitter which is reflected back and is detected by the IR detector. The sensor carry out triangulation for distance measuring, therefore it can measure the distances to an obstacles in a range from 4 to 30 cm. The analog outputs of each sensor are converted by the A/D converters of the microcontroller board. The signals are then filtered by low-pass filters and the actual distance is obtained using look-up tables.

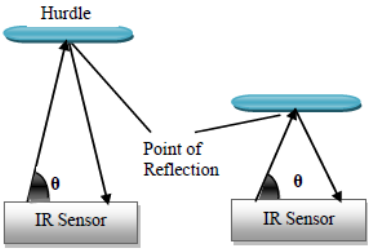

Figure 2. Functioning of an IR sensor.

The Sharp GP2D120 sensor uses an IR emitter and a linear CCD array detector that is 3/4" away from the IR emitter in order to detect the distance from the object.

A picture of the mobile robot is presented in the Fig. 3. The 3 IR proximity sensors are mounted in front of the robot platform, like in Fig. 3 and they are namely *Left_Sensor* (*SL*), *Front_Sensor* (*SF*), *Right_Sensor* (*SR*)..

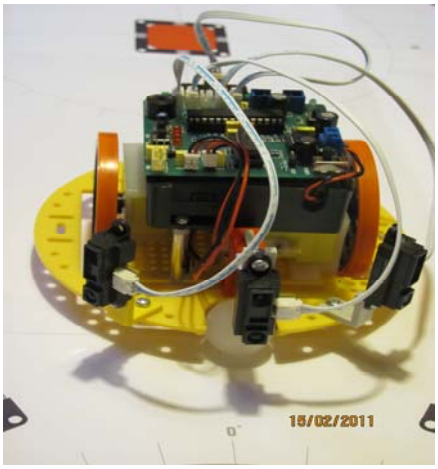

Figure 3. Two wheels differential robot with IR sensors.

Each GP2D120 (IR) sensor is attached at 60° on the left and right side measure from the central axis, like in Fig. 3.

The mechanical structure of this robot has a platform on which are mounted two DC motors. The motors control through gears the differential wheels. This is the most common configuration for a differential drive which use two independent drive motors, one for each side wheel. Also a castor wheel at the rear of the platform to stabilize the vehicle is used. The speed and direction of each wheel will determined the overall direction and speed of the robot. The speed control of

each wheel is made by sending PWM signals to the motor controller, and the motor controller receives PWM signals from the microcontroller board.

3. FUZZY LOGIC CONTROLLER

Fuzzy logic unlike classical logic is tolerant to imprecision, uncertainty and partial truth. In the context of mobile robot control, a fuzzy logic based system has the advantage that it allows intuitive nature of sensor-based navigation and can easily transform linguistic information into control signals [14].

The basic structure of a fuzzy logic controller (FLC) consists of three conceptual components:

1) fuzzification of the input–output variables;
2) rule base that contains a set of fuzzy rules;
3) reasoning mechanism that performs the inference procedure on the rules and given facts to derive a reasonable output.

A general structure for a fuzzy inference system with crisp output is shown in Fig. 4 [16].

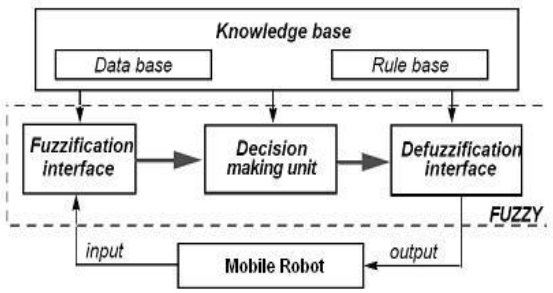

Figure 4. Configuration of a fuzzy logic controller.

The fuzzifier maps crisp input into a fuzzy set, which is input for the inference engine. The inference engine combines rules and gives mapping from fuzzy input sets for fuzzy output sets. The defuzzifier produces crisp output from a fuzzy set that is output for the inference engine. The most common defuzzification approaches include centroid, maximum-decomposition, center of maxima and height.

A. *Fuzzy Sets for Sensors*

The triangular membership functions are used for their simplicity. It means the sum of the membership functions for any variables at any given point in the domain is equal to *one*. In a fuzzy set **S**, each element *x* of the set is assigned with a degree of membership in $\mu_S(x): \mathrm{R} \to [0 \quad 1]$, which is measured by a membership function $\mu_S(x)$. The membership function is *zero* when *x* does not belong to **S** at all, *one* when *x* belongs to totally to **S** and $0 < \mu_S(\mathrm{x}) < 1$ when belongs to partially to **S** [4], [6].

At first we have considered the 3 membership functions: μ_N – near, μ_F – far, μ_M – mediun for each sensor. But it gives us a number of 27 rules for the inference system.

The Matlab graphical representation of these situations for the IR sensors is given by Fig. 5.

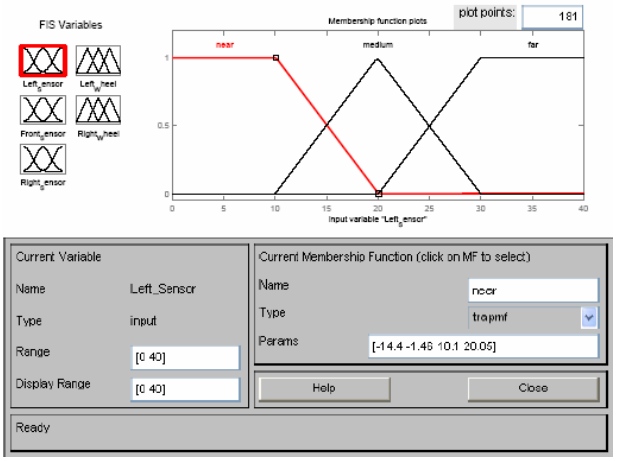

Figure 5. Membership functions $\mu_S(\mathrm{x})$ for the IR sensor.

The distance of area covered by each sensor can be set up on four level distances: 10, 20, 30 or more than 30 cm, respectively. The levels for each sensor are presented in the Table I. It can be seen that the distances measured by each sensor (*left, front* and *right*) are setup on four levels: 10cm, 20cm, 30cm or more than 30 cm, respectively. The levels for each sensor are presented in the Table I.

TABLE I
MEMBERSHIP FUNCTIONS FOR SENSORS

Crisp_Value	Near	Far	DetectC_var	Distance
0	0	1	0	> 30 cm
1	0	0,66	1	30 cm
2	0	0,33	2	20 cm
3	1	0	3	10 cm

Under the distance of 10 cm the obstacle is considered to be *very near* from the robot. The crisp values associated with the distance to the object detected by the sensors are defined by first column of Table I.

B. *Fuzzy Sets for Motors*

The control problem for the two-wheel mobile robot is how to independently control the left-motor and right-motor. Two fuzzy commands are used to control the left wheel and right wheel, respectively. Each wheel can be controlled to move *forward*, *stop* and *reverse*. Combined these commands for each wheel, the robot movement can be described by the following 7 fuzzy sets: (*go_forward*, *go_reverse*, *turn_on_right*, *turn_on_left*, *quick_turn_left*, *quick_turn_right, stop*), as is presented by Table II.

TABLE II
TYPE OF ROBOT MOVEMENT

Robot Movement	Left_Motor	Right_Motor
Go_Forward	Forward_Medium	Reverse_Medium
Go_Reverse	Reverse_Medium	Forward_Medium
Turn_on_left	Forward_Fast	Stop
Turn_on_right	Stop	Reverse_Fast

Quick_turn_left	Forward_Fast	Reverse_Fast
Quick_turn_right	Reverse_Fast	Forward_Fast
Stop	Stop	Stop

For each wheel we have introduced three membership functions (*left stop* and *right*) namely: μ_{FWD}, μ_{REV} and μ_{STOP}. Each of these membership functions are presented in Fig. 6 (μ_{LFWD}, μ_{LSTOP} and μ_{LREV}) for the *left_motor* and in Fig. 7 (μ_{RFWD}, μ_{RSTOP} and μ_{RREV}) for the *right_motor*.

When the robot is moving and the sensors detect an obstacle, a reactive strategy will adjust de robot movement by controlling the speed of wheels. This behavior based reactive model is widely used for robots operating in uncertain dynamic environments, combining information from many sensors [1], [9], [16].

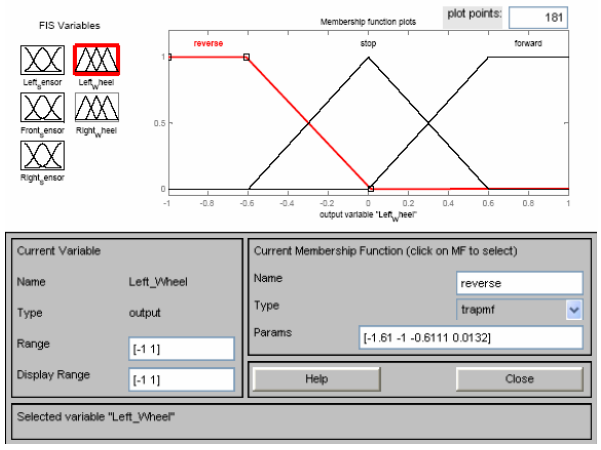

Figure 6. Left wheel membership functions – *forward, stop* and *reverse* (μ_{LFWD}, μ_{LSTOP} and μ_{LREV}).

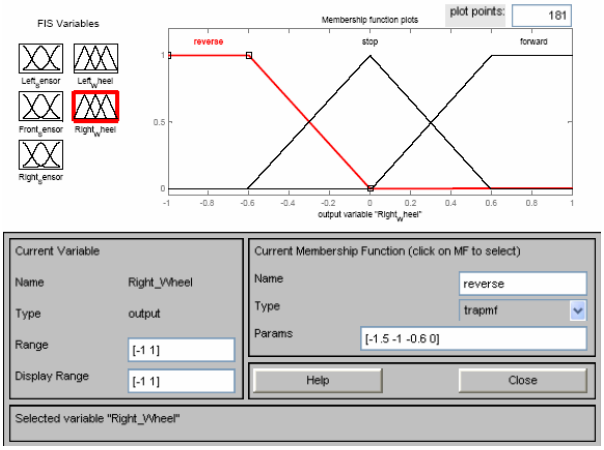

Figure 7. Right wheel membership functions – *forward* and *reverse* (μ_{RFWD}, μ_{RSTOP} and μ_{RREV}).

The control structure is based on a task for avoiding obstacles; the input for the control system is data from sensors and the outputs are the motor commands, as it is shown in Fig. 8. The mobile robot wheels' are independently controlled.

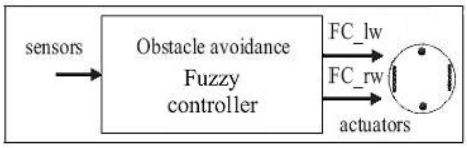

Figure 8. Control structure for avoiding obstacles.

C. *Fuzzy Inference System*

The controller must have 3 inputs and 2 outputs. The inputs are the distance between the robot and the obstacle detected and the outputs are the motors speed. For every input, we have defined 3 membership function (*far, medium* and *near*) and for each output we have defined 3 membership functions (*Reverse, Stop* and *Forward*). It means that there are 216 $(2^3 \cdot 3^3)$ possible rules. The minimum number of necessary rules is three but of course, the obtained behavior is very primitive one. Another important factor is the computational time and this was one of the main reasons that led us to simplify the fuzzy algorithm. In order to reduce the number of the rule we consider only two membership function for the input IR sensors (*far, near*). Analyzing the rules and after some experimental testing we decided to take only 8 rules. The fuzzy control rules for avoiding collision with any obstacles are given in the Table III.

The rules are presented below.

RULE 1
IF (LEFT_SENSOR_L IS FAR) AND
(FRONT_SENSOR IS FAR) AND (RIGHT_SENSOR IS FAR)
THEN (LEFT_MOTOR IS.FWD) (MOTOR_R IS FWD)

RULE 2
IF (LEFT_SENSOR IS FAR) AND
(RIGHT_SENSOR IS FAR) AND (RIGHT_SENSOR IS NEAR)
THEN (LEFT_MOTOR IS REV) (MOTOR_R IS FWD)

RULE 3
IF (LEFT_SENSOR_L IS FAR) AND
(FRONT_SENSOR IS NEAR) AND (RIGHT_SENSOR IS NEAR)
THEN (LEFT_MOTOR IS.REV) (MOTOR_R IS FWD)

RULE 4
IF (LEFT_SENSOR IS FAR) AND
(RIGHT_SENSOR IS NEAR) AND (RIGHT_SENSOR IS FAR)
THEN (LEFT_MOTOR IS REV) (MOTOR_R IS FWD)

RULE 5
IF (LEFT_SENSOR_L IS NEAR) AND
(FRONT_SENSOR IS NEAR) AND (RIGHT_SENSOR IS FAR)
THEN (LEFT_MOTOR IS.FWD) (MOTOR_R IS REV)

RULE 6
IF (LEFT_SENSOR IS NEAR) AND
(RIGHT_SENSOR IS NEAR) AND (RIGHT_SENSOR IS NEAR)
THEN (LEFT_MOTOR IS FWD) (MOTOR_R IS REV)

RULE 7
IF (LEFT_SENSOR_L IS NEAR) AND
(FRONT_SENSOR IS FAR) AND (RIGHT_SENSOR IS NEAR)
THEN (LEFT_MOTOR IS.FWD) (MOTOR_R IS REV)

RULE 8
IF (LEFT_SENSOR IS NEAR) AND
(RIGHT_SENSOR IS FAR) AND (RIGHT_SENSOR IS FAR) THEN
(LEFT_MOTOR IS FWD) (MOTOR_R IS REV)

TABLE III
FUZZY RULES FOR THE ROBOT MOVEMENT

Left_Sensor	Frot_Sensor	Right_Sensor	Left_Motor	Right_Motor
Far	Far	Far	Forward	Forward
Far	Far	Near	Reverse	Forward
Far	Near	Near	Reverse	Forward
Far	Near	Far	Forward	Reverse
Near	Near	Far	Forward	Reverse
Near	Near	Near	Reverse	Reverse
Near	Far	Near	Forward	Reverse
Near	Far	Far	Forward	Reverse

All of these rules must act together so as the robot to avoid any obstacle meets on his way.

D. Defuzzification

Defuzzification is an operation with the aim to produce a non-fuzzy control action. It transforms fuzzy sets into a crisp value.

Among many defuzzification methods, the center of gravity method was chosen. The method generates the centre of area of the resulting fuzzy set of a control action.

The following examples will present how this method works for two different cases of the data sensors. In the first case a combination of $[1 \quad 0 \quad 0]$ was considered for the IR sensors. It means that *left_sensor* detects an obstacle at a *far* distance (more than 30 cm) while the *front_sensor* and *right_sensor* respectively, doesn't detect any object. Figures 9*a* and 9*b* show the robot position in such a case and what are the motor commands obtained by applying the center of gravity defuzzyfication method.

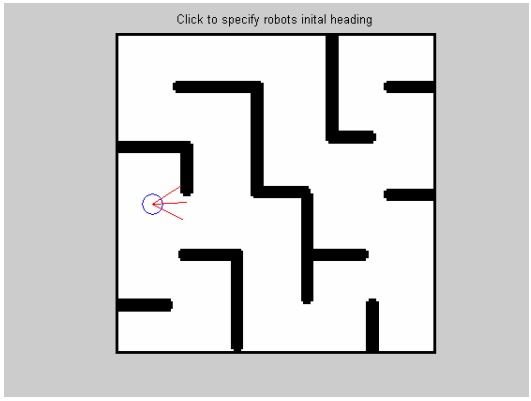

Figure 9a. Robot position for a $[1 \quad 0 \quad 0]$ IR sensors combination.

The second scenarios is for a combination of sensors equal to $[2 \quad 0 \quad 0]$ which means that *left_sensor* detects an obstacle at a *medium* distance (10cm < 20 cm < 30cm) while the *front_sensor* and *right_sensor* respectively, do not detect any object. Fig. 10a presents such a situation for the robot while the Fig. 10b shows the *left* and *right* motor commands resulted by defuzzification.

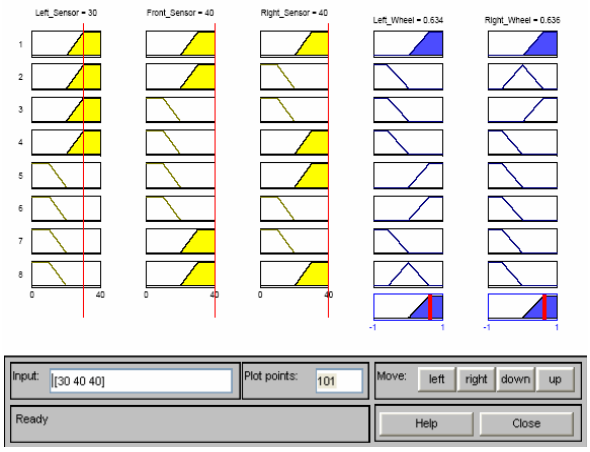

Figure 9b. Crisp value of commands for an input sensor i=[1 0 0]

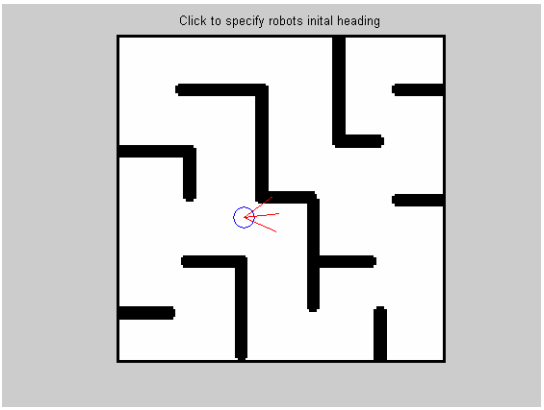

Figure 10a. Robot position for a $[2 \quad 0 \quad 0]$ sensor combination.

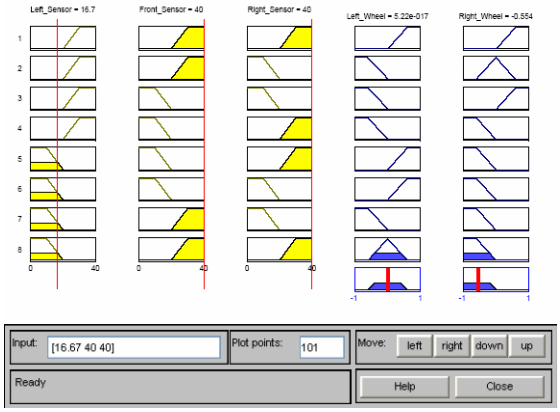

Figure 10b. Crisp value of commands for an input sensor i=[2 0 0].

4. SIMULATION AND EXPERIMENTAL RESULTS

Autonomous navigation is related to the ability of a mobile robot to move in an unknown environment without hitting any obstacles. For experimental tests a small differential mobile robot presented in Fig. 3 was used. The most of navigation scenarios cover aspects including, *wall following*, *collision avoidance*, *corridor following* or *goal seeking*. Based on robot kinematics model we have simulated and analyzed the robot behavior under Matlab/Simulink. Such a simulated scenario, where the robot path and the range finder of each sensor is shown in Fig. 11.

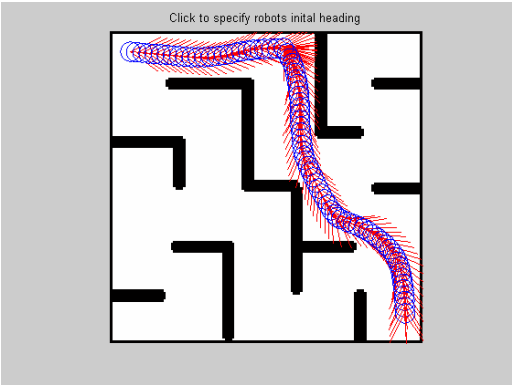

Figure 11. Free collision navigation.

The fuzzy reactive controller was implemented on an Arduino compatible hardware prototyping platform, and the block scheme is depicted in of Fig. 12.

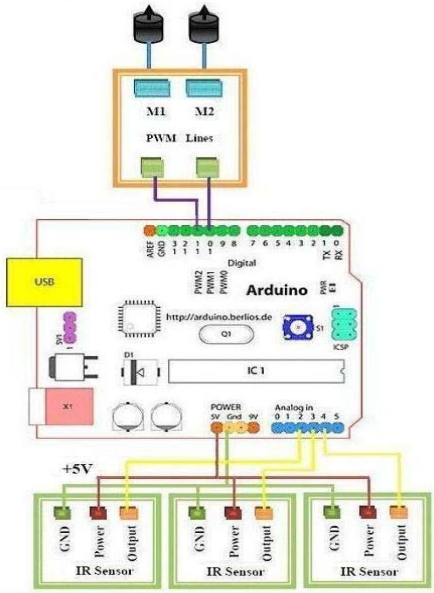

Figure 12. Mobile robot hardware prototyping platform.

Each independent wheel of robot is driven by geared PM DC motor. The motors are controlled independently by the PWM outputs of the microcontroller. The Sharp GP2D120 IR sensors mounted on the front of the robot are connected to the analog input port of the microcontroller. The robot behavior was tested to perform *collision-free* navigation and *corridor following*. The experimental tests have shown that proposed fuzzy reactive controller has a short reaction time and a rapid decision-making of the obstacle avoidance process.

CONCLUSION

In this paper, we have developed and tested a sensor-based methodology for robot navigation in unknown environments where a fuzzy logic controller translates the sensor measurements directly to actuator actions. The controller was simulated and tested on a small differential mobile robot, to perform *collision-free* and *corridor following* navigation. The experimental results have shown that the proposed architecture provides an efficient and flexible solution for the light autonomous differential mobile robots. Learning methods for fuzzy controller can improve the robot behavior.

REFERENCES

[1] Ronald C. Arkin; Behaviour-Based Robotics, The MIT Press, Cambridge, 1998.
[2] G W. Lucas; A Tutorial and Elementary Trajectory Model for the Differential Steering System of Robot Wheel Actuators, The Rossum Project, 2001
[3] G. Dudek, M. Jenkin; Computational Principles of Mobile Robotics, Cambridge University Press, New York, 2000
[4] Babuska, Robert; - Fuzzy Logic for Engineering Applications (et 29-30), Control Engineering Laboratory, Faculty of Information Technology and Systemş Delft University of Technology, Delft, the Netherlands, March 3, 1999
[5] Ibrahim, Ahmed; Fuzzy Logic for Embedded System Applications, Elsevir Science, U.S.A., 2004
[6] Mendel, M., Jerry; Fuzzy Logic Systems for Engineering, A Tutorial IEEE Progeedings of the IEEE, Vol. 83, no.3, March 1993
[7] Jantzen, Jan; Design of Fuzzy Controllers, Technical University of Denmark, Dept. of Automation, Lyngby, Denmark, Technical Report 98-E-864, 19 Aug. 1998
[8] K-team S.A; Khepera User Manual version 5.0, Lausanne, 1998.
[9] R. Brooks, "A robust layered control system for a mobile robot," *IEEE Trans. Robotics Automat.*, vol. RA-2, pp. 14–23, Feb. 1986.
[10] J. S. R. Jang, C. T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*.Englewood Cliffs, NJ: Prentice-Hall, 1997.
[11] S.-B. Cho and S.-I. Lee, "Mobile robot learning by evolution of fuzzy for a miniature mobile robot," in *Proc. Knowledge-Based Electron. Syst.*,
[12] Saffiotti, Alessandro; Ruspini Enrique H.; Konolige, Kurt– Using Fuzzy Logic for Mobile Robot Control, Chapter 5 of the International Handbook of Fuzzy Sets, D. Dubois, H. Prade, and H. Zimmermann, editors, Kluwer Academic Publisher, forthcoming in 1999
[13] Surmann, Hartmut; Huser, Jörg; Peters, Huser and Liliane - A Fuzzy System for Indoor Mobile Robot Navigation, Proc. of the Fourth IEEE Int. Conf. on Fuzzy Systems, pp. 83 - 86, 20 - 24.03.1995, Yokohama, Japan
[14] Tunstel, Edward; Lippincott, Tanya; Jamshidi, MO - Introduction to Fuzzy Logic Control With Application to Mobile Robotics, NASA Center for Autonomous Control Engineering Department of Electrical and Computer Engineering, 1996
[15] Vamsi Mohan Peri; Fuzzy Logic Controller for an Autonomous Mobile Robot, Electrical and Electronics Engineering, Jawaharlal Nehru Technological University, India, May, 2002.
[16] Jelena Godjevac; Comparative Study of Fuzzy Control, Neural Network Control and Neuro-Fuzzy Control, Technical report 103/95, Feb. 1995, EPFL, Lausanne, Swiss

Steady State and Dynamic Behavior of a Permanent Magnet Flux-Switching Machine

SOMEŞAN Liviu[1], PĂDURARIU Emil[1], VIOREL Ioan-Adrian[1], SZABO Lorand[1]
[1]Department of Electrical Machines and Drives,
Technical University of Cluj-Napoca, Faculty of Electrical Engineering,
28 Memorandumului Street, 400114 Cluj-Napoca, e-mail: liviu.somesan@mae.utcluj.ro

Abstract *– The permanent magnet flux-switching machine (PMFSM) is a new brushless machine with the magnets in the stator, offering the advantages of robust rotor structure, high power density and high efficiency. In this paper a typical structure of a PMFSM with 12 stator poles and 10 rotor poles is considered and its performance are checked via a two dimensions finite element method (2D-FEM) analysis. The PMFSM's dynamic behaviour is studied too and the conclusions in the case of the PMFSM sample end the paper.*

Keywords: *permanent magnet flux-switching machine, dynamic behavior.*

I. INTRODUCTION

Permanent magnet flux-switching machine (PMFSM) is a relatively new category of electric machines even it was introduced, in a basic structure, almost half a century ago as a one phase machine. The basic model of PMFSM was described in [1], where Rauch and Johnson proposed a new type of motor with permanent magnets placed in the stator in order to better control their temperature. It was brought back to the scene [2] due to a multitude of reasons, including the extended limit of permanent magnetic materials and the possibility of sophisticated computer-aided motor design.

PMFSM has a unique structure, excellent performance and wide constant power speed range. Therefore, it has recently been the subject of considerable research thanks to the advantages of high power density, high efficiency, mechanical robustness and torque capability [3-6]. Furthermore, the PMFSM can be used with success in harsh operating environments, such as aerospace, automotive and wind energy applications [7, 8].

The PMFSM is very similar to the doubly salient permanent magnet (DSPM) machine and come out of the flux reversal machine (FRM) [4, 5].

This paper takes into consideration a typical three phase structure of a permanent magnet flux-switching machine with 12 stator poles and 10 rotor poles.

In the second part, the particular PMFSM is described briefly and some hints on the machine simplified model and its sizing-designing algorithm are given.

In the third part the sizing-designing algorithm is validated by the results obtained via a two dimensions finite element method (2D-FEM) analysis done for a sample machine operating as a motor. The results obtained via simulations can provide valuable information on the machine's behavior.

Finally the PMFSM's dynamic behavior is studied using the flux oriented control.

The conclusions and the final considerations in the case of the sample of PMFSM end the paper

II. PMFSM STRUCTURE AND SIMPLIFIED MODEL

Fig.1 shows a typical three phase PMFSM structure. As it can be seen, the rotor of the machine is similar to that of a switched reluctance motor, the number of rotor poles and stator poles differing by two, 10 rotor poles and 12 stator poles. Also the concentrated windings employed in the PMFSM's are similar to those in the switched reluctance motor (SRM).

The only difference compared to the SRM's consists of the configuration of the stator which contains 12 segments of "U" shape magnetic cores, between which 12 pieces of permanent magnets are inset, the direction of magnetization being reversed from one permanent magnet to the following.

The stator winding comprises concentrated coils, each coil being wound around a pole which contains two adjacent laminated segments and a permanent magnet. This fact leads to low copper loss due to the short end-windings.

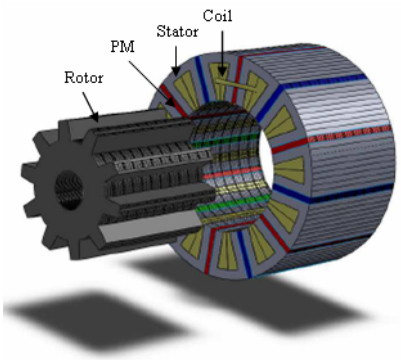

Figure 1. A 3-phase 12/10 PMFSM prototype.

The developed sizing-design procedure is presented fully in the flow chart given in Fig. 2. The basic flux calculation is done by employing simplified magnetic equivalent circuits (MEC), as is the one given in Fig. 3. The magnetic equivalent circuits are constructed for different situations, in order to calculate the no-load main flux, the armature reaction and the most important leakage fluxes.

For the aligned position, the MEC is presented in Fig. 3 in case of no-load fluxes, when there is no current in the armature winding. Φ_g Φ_{PM} ,$\Phi_{\sigma PMx}$ are no-load air-gap main flux, permanent magnet flux and permanent magnet leakege flux, repesctively R_{PM}, $R_{\sigma PM}$ are reluctance corresponding to the Φ_{PM} ,$\Phi_{\sigma PMx}$ fluxes and R_g, R_{Fet}, R_{FepR} are the air-gap, stator tooth and rotor pole reluctances.

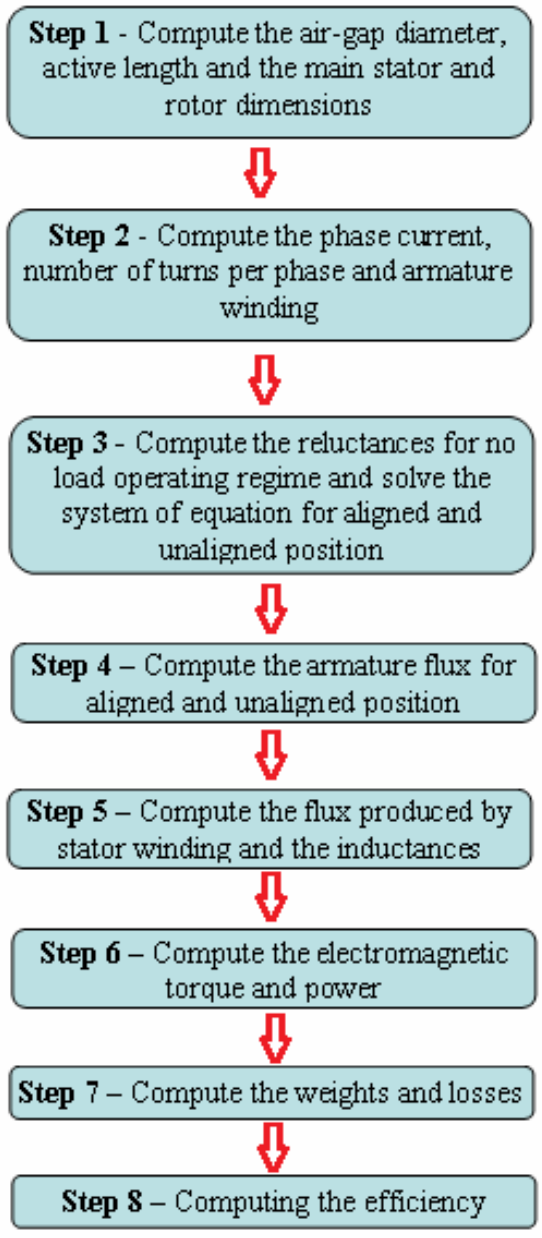

Figure 2. Specific PMFSM sizing-design flow chart.

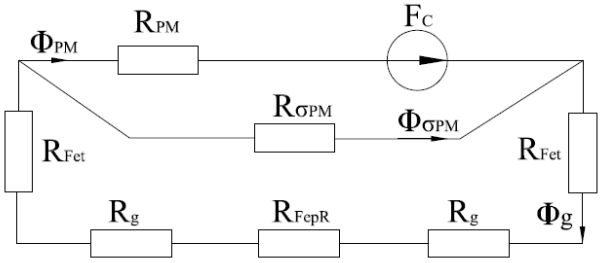

Figure 3. No-load MEC, unaligned rotor position.

While the reluctances of the ferromagnetic regions and the permanent magnets are relatively easy to determine, the reluctances of the air-gap region R_g are more complicated. The air-g permeance is:

$$P_g = \frac{\mu_0 \cdot b_{pR} \cdot L_{st}}{2 \cdot g} + 0.26 \cdot \mu_0 \cdot L_{st} + \frac{\mu_0 \cdot L_{st}}{\pi} \cdot \ln\left[1 + \pi \cdot \frac{h_{PM} - h_{yS}}{2 \cdot g}\right] \quad (1)$$

In P_g is the air-gap permeance, b_{pR} is the rotor pole width, $L_{st,}$ is the stack length, g is the air-gap length, h_{PM} is the permanent magnet height, h_{yS} is the stator yoke height and μ_0 is the free space permeability.

It results the air-gap reluctance:

$$R_g = \frac{1}{P_g} \quad (2)$$

The main designing specifications for the sample PMFSM are:

- Rated output power P_{out} = 30 kW
- Rated phase voltage U_f = 230 V
- Rated speed n = 3000 rpm

The main dimensions of the flux switching machine calculated using the algorithm presented here are evinced in table I.

Table I. Main geometric dimensions and parameters of PMFSM

Number of rotor poles, $\mathbf{Q_R}$	-	10
Number of stator poles, $\mathbf{Q_S}$	-	12
Machine‘s outer diameter, $\mathbf{D_{out}}$	m	0,277
Shaft diameter, **dax**	m	0,045
Air-gap diameter, $\mathbf{D_g}$	m	0,159
Air-gap length, **g**	m	0,0007
Stator PM height, $\mathbf{b_{PM}}$	m	0,008538
Rotor pole pitch, $\boldsymbol{\tau_R}$	m	0,0497
Rotor pole width, $\mathbf{b_{pR}}$	m	0,01457
Stator pole pitch, $\boldsymbol{\tau_S}$	m	0,041832
Stator pole width, $\mathbf{b_{pS}}$	m	0,011418
Stator slot height, $\mathbf{h_{slS}}$	m	0,04612
Rotor yoke height, $\mathbf{h_{yR}}$	m	0,03475
Stack length, $\mathbf{l_{st}}$	m	0,159
Stator yoke height, $\mathbf{h_{yS}}$	m	0,01249
Number of turns per phase, $\mathbf{N_t}$	-	36
Phase current, $\mathbf{I_f}$	A	58

III. 2D-FEM ANALYSIS

The 2D-FEM analysis by using the Cedrat FLUX 2D environment was employed in order to check the analytically obtained results via the sizing-designing calculation.

For the proposed PMFSM the calculated values of the phase back-**emf** waveforms at 3000 rpm is illustrated in Fig. 4, which shows that the predicted value from the analytical model is in good agreement with the FEM calculated one. Essentially, the results show that the back-**emf** is almost to sinusoidal with negligible harmonic distortion indicating that the proposed PMFSM is inherently suitable for brushless AC operation, that is, sine-wave operation.

The 2D-FEM computed electromagnetic torque and cogging torque are displayed in Figs. 5 and 6 versus rotor position. The rotor was moved over a complete electrical period with an increment of 1 mechanical degree, one electrical period corresponding to 36 mechanical degrees.

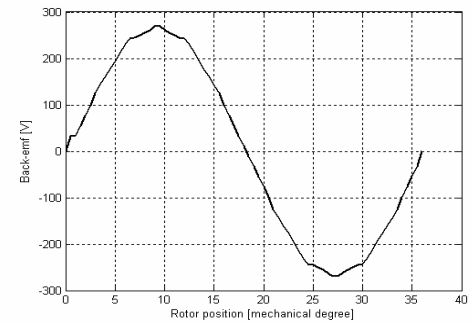

Figure 4. Phase back-emf waveform of the PMFSM.

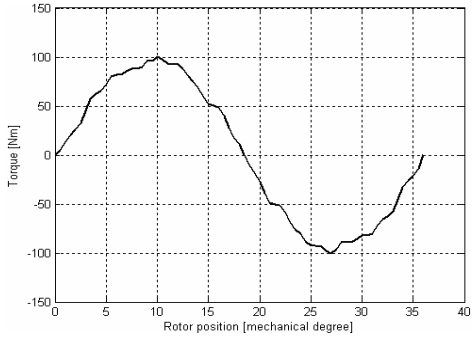

Figure 5. Electromagnetic torque of the PMFSM.

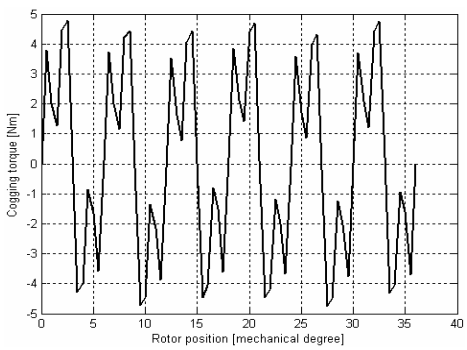

Figure 6. Cogging torque of the PMFSM.

IV. DYNAMIC SIMULATION OF THE PMFSM

The permanent magnet flux-switching machine behavior, in a dynamic regime, can be obtained by using the following equations:

$$\begin{aligned} v &= R \cdot i + \frac{d\psi}{dt} \\ \psi &= f(i,\theta) \\ \theta &= \omega \cdot t \\ T &= \frac{J}{p} \cdot \frac{d\omega}{dt} + T_l \\ T &= i \cdot \frac{\partial \psi}{\partial \theta} \cdot p \end{aligned} \tag{3}$$

where v is the terminal phase voltage, i is the phase current, R is phase resistance, Ψ is the phase flux linkage, θ is the rotor angular position in electrical radians, ω is the electric angular frequency, p is the number of pole pairs and T, T_l are the electromagnetic, respectively the load torque.

Since a three phase PMFSM behaves exactly the same way as a synchronous machine, it is possible to obtain for such a machine a *dq0* model in which the parameters are the same as for the original machine.

The permanent magnet excited flux switching machine equations in a *dq0* model, in a reference frame fixed in the rotor, are:

$$\begin{aligned} v_d &= R \cdot i_d + \frac{\partial \psi_d}{\partial t} - \omega \cdot \psi_d \\ v_q &= R \cdot i_q + \frac{\partial \psi_q}{\partial t} + \omega \cdot \psi_q \\ \psi_d &= \left(L_{s\sigma} + M_d\right) \cdot i_d + \psi_{Em} \\ \psi_q &= \left(L_{s\sigma} + M_q\right) \cdot i_q \\ T &= \frac{3}{2} Q_R \cdot \left(\psi_m \cdot i_q + \left(L_d \cdot L_q\right) \cdot i_d \cdot i_q\right) \end{aligned} \tag{4}$$

where v_d, v_q are the terminal *d*- and *q*-axis voltages, i_d, i_q are the *d*- and *q*-axis currents, Ψ_d, Ψ_q are the stator *d*- and *q*-axis flux linkages, Ψ_{EM} is the permanent magnet, or no load flux linkage, $L_{s\sigma}$ is the stator phase leakage inductance and M_d, M_q are the *d*- and *q*-axis main inductance.

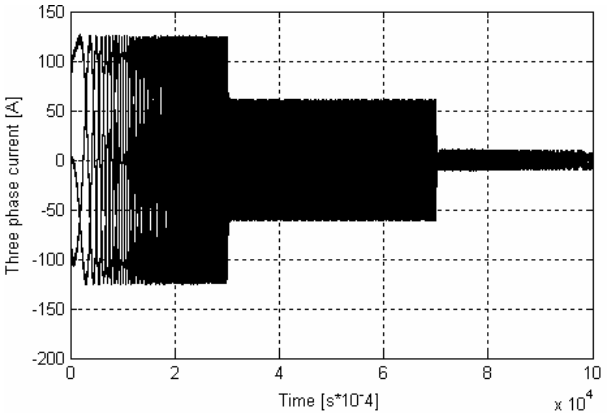

Figure 7. Three phase currents.

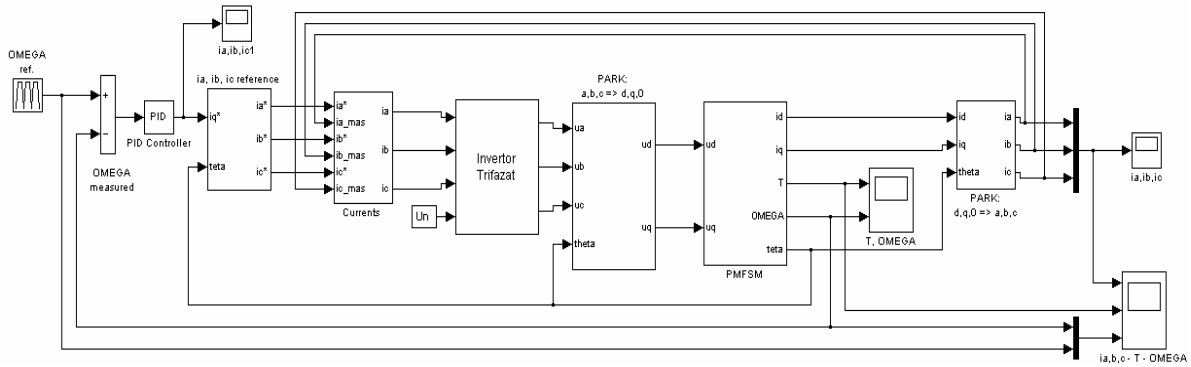

Figure 8. SIMULINK model of the PMTFM.

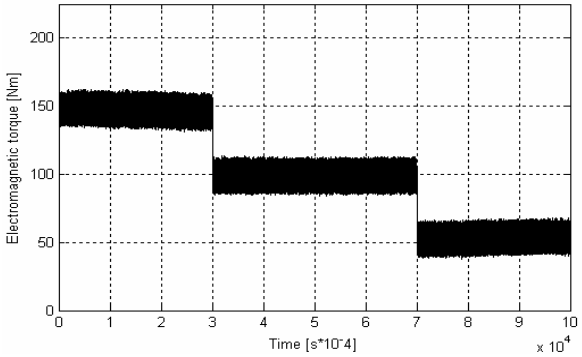

Figure 9. Electromagnetic torque.

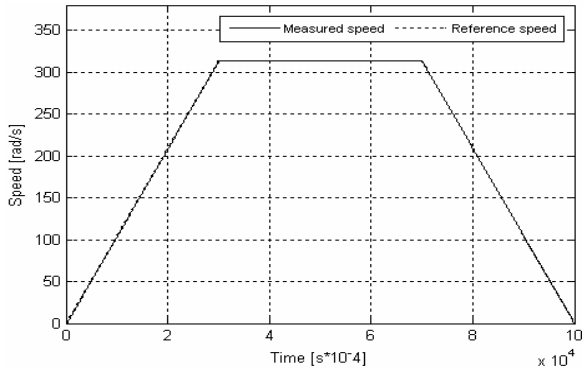

Figure 10. Reference speed versus measured speed.

The Matlab/SIMULINK model for the PMFSM is presented in Fig. 8. The model is based on the flux oriented control and it contains the reference speed measured in rad/s, a PI controller, the reference currents and the PMFSM *dq0* model.

The load torque was imposed at 10

0.8 Nm and it remains at this value for the entire simulation.

The measured currents, electromagnetic torque and speed versus time are shown in Figs. 7, 9 and 10, respectively.

V. CONCLUSIONS

In this paper a typical structure of permanent magnet flux-switching machine (PMFSM) with 12 stator poles and 10 rotor poles was proposed and its performance were checked via a dimension magnetic field calculation (2D-FEM) by using the Cedrat FLUX 2D environment.

The results obtained via simulations provide valuable information on the machine's behavior.

The PMFSM's dynamic behavior has been studied too, using the flux oriented control.

ACKNOWLEDGEMENTS

This paper was supported by the project "Doctoral studies in engineering sciences for developing the knowledge based society "SIDOC" contract no. POSDRU/88/1.5/S/60078, project co-funded from European Social Fund through Sectorial Operational Program Human Resources 2007-2013.

REFERENCES

[1] S. E. Rauch and L. J. Johnson, "Design principles of flux-switching alternators," *AIEE Trans.* Vol. *74III*, pp. 1261-1268, 1955.

[2] B. Sarlioglu, Y. F. Zhao, and T. A. Lipo, "A novel doubly saliency single phase permanent magnet generator," in *Proc. IEEE Industry Applications Soc. Annu. Meeting*, vol. 1, 1994, pp. 9–15.

[3] E. Hoang, A. H. Ben-Ahmed and J. Lucidarme, "Switching flux permanent magnet polyphased synchronous machines," *Proc. 7th Eur. Conf. Power Electron. and Applicat.*, vol. 3, pp. 903-908, 1997.

[4] I. Boldea, C. Wang and S. A. Nasar, "Design of a three-phase flux reversal machine", *EMPS,* vol. 27, pp. 849-863, 1999.

[5] R. P. Deodhar, S. Andersson, I. Boldea and T. J. E. Miller, "The flux reversal machine: A new brushless doubly salient permanent magnet machine", *Proc. Of IEEE-IAS Annual Meeting, pp. 786-793, 1996.*

[6] G. Henneberger and I.A. Viorel, "Variable reluctance electrical machines", *Shaker Verlag, Aachen, Germany, 2001.*

[7] A. S. Thomas, Z. Q. Zhu, R. L. Owen, G. W. Jewell and D. Howe, "Multiphase Flux-Switching Permanent-Magnet Brushless Machine for Aerospace Application", *IEEE Trans. Ind.Appl.,* vol. 45, No. 6, pp. 1971-1981, November/December 2009.

[8] Z. X. Fang, Y. Wang, J.X Shen and Z. W. Huang, "Design and analysis of a novel flux-switching permanent magnet integrated starter-generator" , IET 4th International Conference on Power Electronics, Machines & Drives PEMD 2008, York, UK, 2008.

Application of Non-Iterative Method in Image Deblurring

STOJANOVIC Igor[1], STANIMIROVIC Predrag[2], MILADINOVIC Marko[2], STOJANOVIC Dragana[1]

[1] 'Goce Delcev' University - Stip, Macedonia,
Department of Signal Processing and Communication Systems Group, Faculty of Computer Science,
Toso Arsov 14, 2000 Stip, Macedonia, E-Mail: igor.stojanovik@ugd.edu.mk, d.stojanovic79@yahoo.com

[2] University of Nis, Serbia,
Department of Computer Science, Faculty of Sciences and Mathematics,
Visegradska 33, 18000 Nis, Serbia, E-Mail: pecko@pmf.ni.ac.rs, markomiladinovic@gmail.com

Abstract *– This paper presents a non-iterative method that finds application in a broad scientific field such as image deblurring. A method for image deblurring, based on the pseudo-inverse matrix is apply for removal of blurr in an image caused by linear motion. This method assumes that linear motion corresponds to an integral number of pixels. Compared to other classical methods, this method attains higher values of the Improvement in Signal to Noise Ratio (ISNR) parameter and of the Peak Signal-to-Noise Ratio (PSNR). We give an implementation in the MATLAB programming package.*

Keywords: *deblurring; image restoration; matrix equation; pseudoinverse.*

I. INTRODUCTION

Blurring is a form of bandwidth reduction of an ideal image owing to the imperfect image formation process [1-3]. It can be caused by relative motion between the camera and the original scene, or by an optical system that is out of focus. When aerial photographs are produced for remote sensing purposes, blurs are introduced by atmospheric turbulence, aberrations in the optical system, and relative motion between the camera and the ground. The field of image restoration is concerned with the reconstruction or estimation of the uncorrupted image from a blurred one. In the use of image restoration methods, the characteristics of the degrading system are assumed to be known a priori.

The method, based on pseudoinverse matrix, is applied for the removal of blur in an image caused by linear motion. For comparison, we used two commonly used filters from the collection of least-squares filters, namely Wiener filter and the constrained least-squares filter [2]. Also we used in comparison the iterative nonlinear restoration based on the Lucy-Richardson algorithm [3].

This paper is organized as follows. In the second section we present process of image formation and problem formulation. In Section 3 we describe a method for the restoration of the blurred image. We observe certain enhancement in the parameters: *ISNR*, *MSE* and *PSNR*, compared with other standard methods for image restoration, which is confirmed by the numerical examples reported in the last section.

II. MODELING OF THE PROCESS OF THE IMAGE FORMATION

We assume that the blurring function acts as a convolution kernel or point-spread function $h(n_1, n_2)$ and the image restoration methods that are described here fall under the class of linear spatially invariant restoration filters. It is also assumed that the statistical properties (mean and correlation function) of the image do not change spatially. Under these conditions the restoration process can be carried out by means of a linear filter of which the point-spread function (PSF) is spatially invariant. If we denote by $f(n_1, n_2)$ the desired ideal spatially discrete image that does not contain any blur or noise, then the recorded image $g(n_1, n_2)$ is modeled as [2]:

$$g(n_1, n_2) = h(n_1, n_2) * f(n_1, n_2) = \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{M-1} h(k_1, k_2) f(n_1 - k_1, n_2 - k_2). \quad (1)$$

The objective of the image restoration is to make an estimate $f(n_1, n_2)$ of the ideal image, under the assumption that only the degraded image $g(n_1, n_2)$ and the blurring function $h(n_1, n_2)$ are given. The problem can be summarized as follows: let H be a $m \times n$ real matrix. Equations of the form:

$$g = Hf, g \in \Re^m; f \in \Re^n; H \in \Re^{m \times n} \quad (2)$$

describe an underdetermined system of m simultaneous equations (one for each element of vector g) and $n = m + l - 1$ unknowns (one for each element of vector f). Here the index l indicates horizontal linear motion

blur in pixels. The problem of restoring an image that has been blurred by linear motion, usually results of camera panning or fast object motion can be expressed as, consists of solving the underdetermined system (2). A blurred image can be expressed as:

$$\begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ \vdots \\ g_n \end{bmatrix} = \begin{bmatrix} h_1 & \cdots & h_l & 0 & 0 & 0 & 0 \\ 0 & h_1 & \cdots & h_l & 0 & 0 & 0 \\ 0 & 0 & h_1 & \cdots & h_l & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & h_1 & \cdots & h_l \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_m \end{bmatrix}. \quad (3)$$

The elements of matrix H are defined as: $h_i = 1/l$ for i=1, 2,..., l. The objective is to estimate an original row per row f (contained in the vector f^T), given each row of a blurred g (contained in the vector g^T) and a priori knowledge of the degradation phenomenon H. We define the matrix F as the deterministic original image, its picture elements are F_{ij} for i=1,…, r and for j=1,…, n, the matrix G as the simulated blurred can be calculated as follows:

$$G_{ij} = \frac{1}{l}\sum_{k=0}^{l-1} F_{i,j+k}, i = 1,\cdots,r,\ j = 1,\cdots,m \quad (4)$$

with $n = m + l - 1$, where l is the linear motion blur in pixels. Equation (4) can be written in matrix form of the process of *horizontal* blurring as:

$$G = \left(HF^T\right)^T = FH^T. \quad (5)$$

Since there is an infinite number of exact solutions for f or F in the sense that satisfy the equation $g = Hf$ or $G = FH^T$, an additional criterion that find a sharp restored matrix is required.

The process of blurring with vertical motion is with the form:

$$g = Hf, g \in \Re^m;\ f \in \Re^r;\ H \in \Re^{m\times r} \quad (6)$$

where $r = m + l - 1$, and l is linear vertical motion blur in pixels. The matrix H is Toeplitz matrix as the matrix given in (3), but with other dimensions. The matrix form of the process of *vertical* blurring of the images is:

$$G = HF, G \in \Re^{m\times n};\ H \in \Re^{m\times r};\ F \in \Re^{r\times n}. \quad (7)$$

Let us first consider a case where the blurring of the columns in the image is independent of the blurring of the rows - s*eparable two-dimensional blur*. When this is the case, then there exist two matrices H_c and H_r, such that we can express the relation between the original and blurred images as:

$$\begin{aligned} &G = H_c F H_r{}^T, G \in \Re^{m_1\times m_2};\\ &H_c \in \Re^{m_1\times r};\ F \in \Re^{r\times n};\ H_r \in \Re^{m_2\times n}. \end{aligned} \quad (8)$$

where $n = m_2 + l_1 - 1$, $r = m_1 + l_2 - 1$, l_1 is linear horizontal motion blur in pixels and l_2 is linear vertical motion blur in pixels.

III. METHOD FOR IMAGE DEBLURRING

We will use the following proposition from [5]:

Let $T \in R^{m\times n}, b \in R^m, b \notin \Re(T)$ and we have a relationship $Tx = b$, then we have $T^\dagger b = u$, where u is the minimal norm solution and $T^\dagger$ is the pseudoinverse matrix of T.

Since relation (2) has infinitely many exact solutions for f, we need an additional criterion for finding the necessary vector for restoration. The criterion that we use for the restoration of blurred image is the minimum distance between the measured data:

$$\min\left(\left\|\hat{f} - g\right\|\right) \quad (9)$$

where $\hat{f}$ are the first m elements of the unknown image f, which is necessary to restore, with the following constraint:

$$\|Hf - g\| = 0. \quad (10)$$

Following the above proposal, only one solution of the relation $g = Hf$ minimizes the norm $\|Hf - g\|$. If this solution is marked by $\hat{f}$, then for it is true:

$$\hat{f} = H^\dagger g. \quad (11)$$

Taking into account the relations of horizontal blurring (2) and (5), and relation (11) solution for the restored image is:

$$\hat{F} = G(H^T)^\dagger = G(H^\dagger)^T. \quad (12)$$

In the case of process of *vertical blurring* solution for the restored image, taking into account equations (6), (7) and (11), is:

$$\hat{F} = H^\dagger G. \quad (13)$$

When we have a *separable two-dimensional blurring* process, the restored image is given by:

$$\hat{F} = H_c^\dagger G (H_r^\dagger)^T. \quad (14)$$

IV. EXPERIMENTAL RESULTS

In this section we have tested the method based on pseudoinverse matrix (GIM method) of images and present numerical results and compare with two standard methods for image restoration called least-squares filters: Wiener filter and constrained least-squares filter and the iterative method called Lucy-Richardson algorithm. The experiments have been performed using Matlab programming language on an Intel(R) Core(TM) i5 CPU M430 @ 2.27 GHz 64/32-bit system with 4 GB of RAM memory running on the Windows 7 Ultimate Operating System.

In image restoration the improvement in quality of the restored image over the recorded blurred one is measured by the signal-to-noise ratio (*SNR*) improvement. The *SNR* of the recorded (blurred and noisy) image is defined as follows in decibels [6]:

$$SNR_g = 10\log_{10}\left(\frac{\text{Variance of } f(n_1,n_2)}{\text{Variance of } g(n_1,n_2)-f(n_1,n_2)}\right). \quad (15)$$

The *SNR* of the restored image is similarly defined as:

$$SNR_{\hat{f}} = 10\log_{10}\left(\frac{\text{Variance of } f(n_1,n_2)}{\text{Variance of } \hat{f}(n_1,n_2)-f(n_1,n_2)}\right). \quad (16)$$

Then, the improvement in *SNR* is given by:

$$ISNR = 10\log_{10}\left(\frac{\text{Variance of } g(n_1,n_2)-f(n_1,n_2)}{\text{Variance of } \hat{f}(n_1,n_2)-f(n_1,n_2)}\right) \quad (17)$$

The simplest and most widely used full-reference quality metric is the mean squared error (*MSE*) [6], along with the related quantity of peak signal-to-noise ratio (*PSNR*). The advantages of *MSE* and *PSNR* are that they are very fast and easy to implement. With *PSNR* greater values indicate greater image similarity, while with *MSE* greater values indicate lower image similarity. Below *MSE*, *PSNR* are defined:

$$MSE = \frac{1}{rm}\sum_{i=1}^{r}\sum_{j=1}^{m}\left|f_{i,j}-\hat{f}_{i,j}\right|^2 \quad (18)$$

$$PSNR = 20\log_{10}\left(\frac{MAX}{\sqrt{MSE}}\right)(\text{dB}) \quad (19)$$

where *MAX* is the maximum pixel value.

A. Horizontal Motion

Fig. 1, Original Image, shows a deterministic original standard Matlab image Camera. Fig. 1, Degraded Image, presents the degraded Camera image for *l*=30. Finally, from Fig. 1, GIM Restored Image, Wiener Restored Image, Constrained LS Restored Image and Lucy-Richardson Restored Image, it is clearly seen that the details of the original image have been recovered.

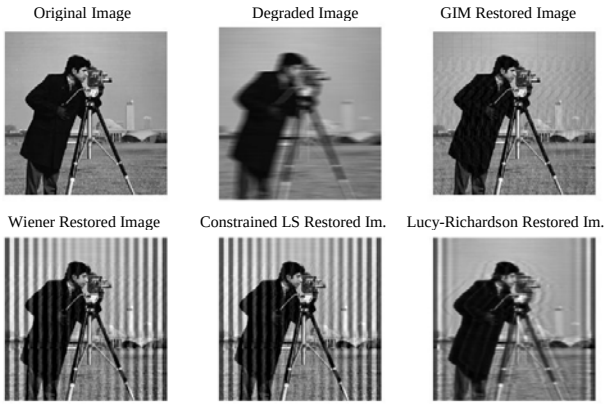

Figure 1. Restoration in simulated degraded Camera image for length of the horizontal blurring process, *l*=30.

The difference in quality of restored images can hardly be seen by human eye. For this reason, the *ISNR* and *MSE* have been chosen in order to compare the restored images. Fig. 2 and Fig. 3 shows the corresponding *ISNR* and *MSE* values. The figures illustrate that the quality of the restoration is as satisfactory as the classical methods or better from them (*l*<100 pixels).

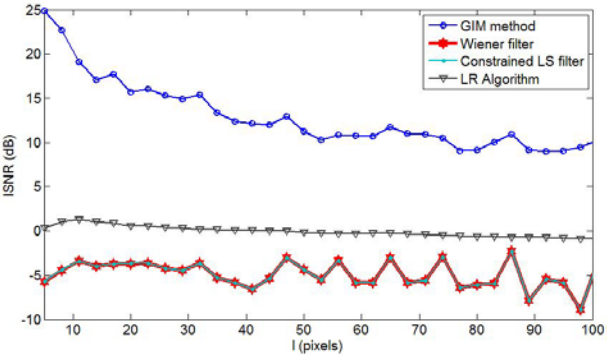

Figure 2. Improvement in signal-to-noise-ratio vs. length of the blurring process in pixels.

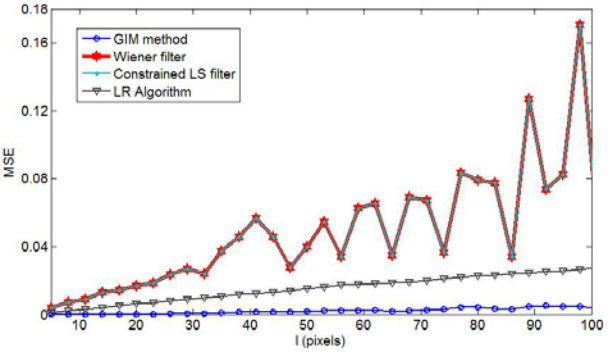

Figure 3. Mean squared error vs. length of the blurring process in pixels.

B. *Vertical Motion*

Obviously the method is not restricted to restoration of images blurered from horizontal motion. The results present in Fig. 4 – 5 refer when we have vertical blurring process.

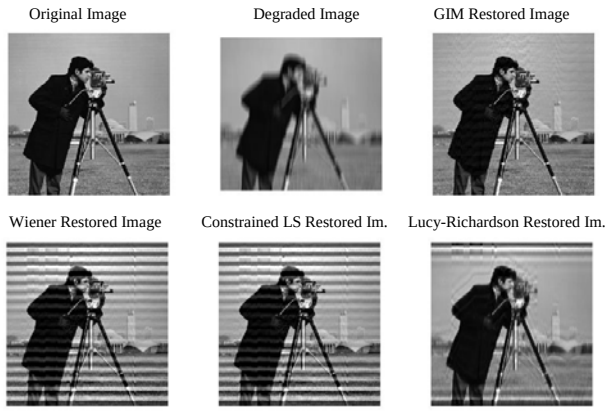

Figure 4. Restoration in simulated vertical degraded image for length of the blurring process, l=30.

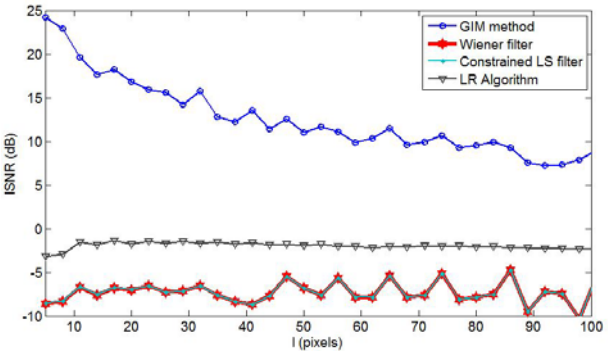

Figure 5. Improvement in signal-to-noise-ratio vs. length of the blurring process in pixels.

C. *Separable two-dimensional blur*

The results for the standard Matlab image Camera in case of separable two-dimensional blur are given on Fig. 6 – 7.

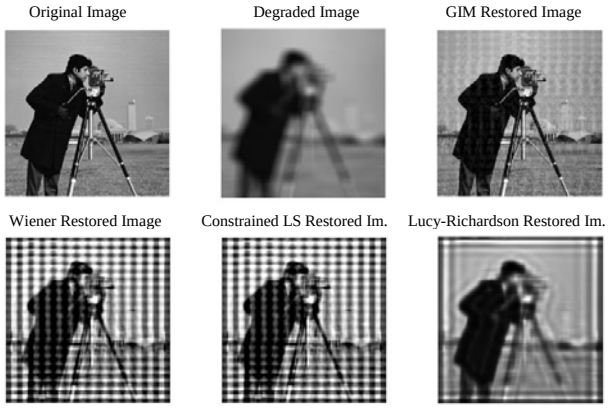

Figure 6. Restoration in simulated degraded Camera image for length of the blurring process l_1=35 and l_2=25.

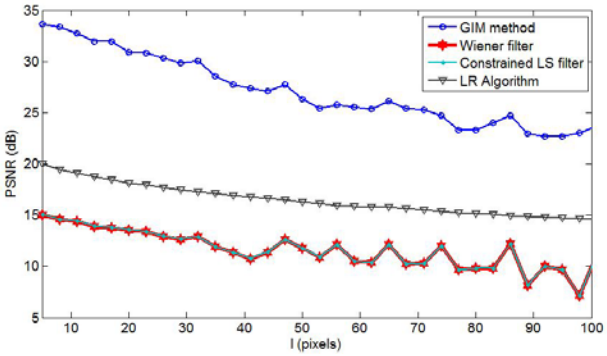

Figure 7. Peak signal-to-noise-ratio vs. length of the blurring process in pixels.

IV. CONCLUSIONS

We introduce a computational method, based on the pseudoinverse matrix, to restore an image that has been blurred by linear motion.

We are motivated by the problem of restoring blurry images via the well-developed mathematical methods and techniques based on pseudoinverse matrix in order to obtain an approximation of the original image.

We present the results by comparing our method and that of the Wiener filter, Constrained least-squares filter and Lucy-Richardson algorithm, well-established restoration methods.

In the method we studied, the resolution of the restored image remains at a very high level, yet the *ISNR* is considerably higher while the computational efficiency is improved in comparison to other methods and techniques.

REFERENCES

[1] J. Biemond, R. L. Lagendijk, and R. M. Mersereau, "*Iterative methods for image deblurring*", Proc. IEEE, 78(5):856–883, 1990.

[2] Al Bovik, *The essential guide to the image processing*, Academic Press, 2009.

[3] Rafael C. Gonzalez, Richard E. Woods, *Digital Image Processing*, 2nd Edition, Prentice-Hall, 2002.

[4] Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins, *Digital Image Processing Using MATLAB*, Prentice-Hall, 2003.

[5] Spiros Chountasis, Vasilios N. Katsikis, and Dimitrios Pappas, "*Applications of the Moore-Penrose Inverse in Digital Image Restoration*", Mathematical Problems in Engineering Volume 2009 (2009).

[6] Ahmet M. Eskicioglu and Paul S. Fisher, "*Image Quality Measures and Their Performance*", IEEE Transactions on Communications, vol. 43, pp. 2959-2965, Dec. 1995.

[7] Zhou Wang and Alan C. Bovik, "*Mean Squared Error: Love It or Leave It? A New Look at Signal Fidelity Measures*", IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, Jan. 2009.

[8] *Image Processing Toolbox User's Guide*: The Math Works, Inc., Natick, MA, 2009.

A Review of Congestion Management Algorithms on Cisco Routers

SZILÁGYI Szabolcs[1], ALMÁSI Béla[2]
[1] Department of Computer Science,
University of Oradea, Faculty of Electrical Engineering and Information Technology,
Universităţii 1, 410087 Oradea, Romania, E-Mail: szilagyi.szabolcs@inf.unideb.hu, sszilagyi@uoradea.ro

[2]Department of Informatics Systems and Networks,
University of Debrecen, Faculty of Informatics,
Kassai 26, 4028 Debrecen, Hungary, E-Mail: almasi.bela@inf.unideb.hu

***Abstract* – This paper presents one of the features of DS (Differentiated Services) architecture, namely the queuing or congestion management. Packets can be put in separate buffer queues, on the basis of DS value. Several forwarding policies can be used to favor high priority packets in different ways. The main reason for queuing is that a router must hold a packet in its memory, and meanwhile the outgoing interface is busy with sending another packet. The queuing tools are covered in the order in which they were added as Cisco IOS features: FIFO (First-In First-Out), CQ (Custom Queuing), PQ (Priority Queuing), WFQ (Weighted Fair Queuing), CBWFQ (Class Based Weighted Fair Queuing) and LLQ (Low Latency Queuing).**

***Keywords:* congestion; queuing; FIFO; PQ; CQ; WFQ; CBWFQ; LLQ.**

I. INTRODUCTION

In the beginning, the Internet was designed for data processing applications where delays were not very important. In the majority of cases a best effort delivery service was enough, and when data was lost or corrupted, the TCP protocol took care of the retransmission and recovery which was necessary. Today these expectations have changed due to the growth of multimedia applications, which require higher bandwidth (they need megabits per second instead of kilobits per second which was used for data processing applications). Nowadays applications are quite sensitive for the delays experienced when transmitting over the Internet. Therefore it is important to keep track of the delay and delay variation (jitter) and ensure they don't overgrow.That is why it is needed to support a variety of traffic with different quality of service (QoS) [1]. The most important side of this is how to share the existing resources while experiencing congestion. In order to proceed with this, it is needed that different mechanisms help differentiate between the types of traffic (prioritize).

The mechanisms which facilitates the queuing on each interface consists in hardware and software components.[2] If the hardware queue, often named "transmit queue" (TxQ) is not congested or full (exhausted), the packets are not kept in the software queue. They are switched directly to the hardware queue, where they are transferred quickly to the medium using FIFO order. In the case when the hardware queue is full, the packets are held in the software queue, processed, and released to the hardware queue based on the software queuing discipline. (Figure 1) [3].The software queuing discipline could be FIFO (First-In First-Out), CQ (Custom Queuing), PQ (Priority Queuing), WFQ (Weighted Fair Queuing), CBWFQ (Class Based Weighted Fair Queuing) and LLQ (Low Latency Queuing).

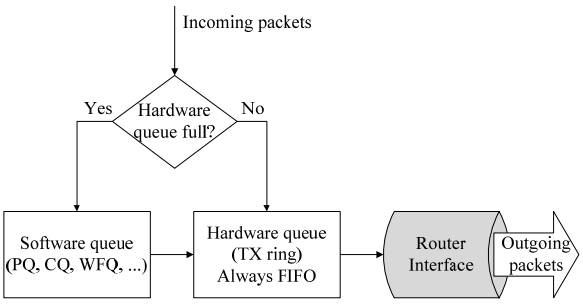

Figure 1. Queuing components.

II. FIFO QUEUING

First-In First-Out (FIFO) is the most simple modality of queuing. The incoming packets are put in a single queue and are processed in the order of receiving them. (Figure 2) Packets are dropped when the FIFO queue is full (tail drop). [2] This queuing type requires little computation and its behavior is very predictable, i.e. the delay of packet is a direct function of the queue size. Cisco IOS defaults to use FIFO on Fast Ethernet and Gigabit Ethernet interfaces with bandwidths above E1 speeds (2.048 Mbps). [4] There are several undesirable characteristics related to this queuing policy, because of its simplistic approach. [1]

- It is not possible to offer different services for different packet classes as all packets are put into the same queue.

- If an incoming flow suddenly becomes bursty, then it is possible for the entire buffer space to be filled by this single flow and other flows will not be serviced until the buffer is emptied.

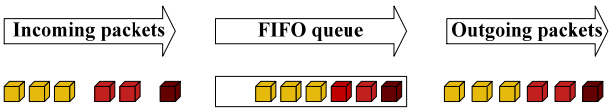

Figure 2. FIFO.

III. PRIORITY QUEUING

"A simple way to offer different services to different classes of packets is Priority Queuing. Its operation involves classifying each incoming packet into different priorities and placing them into separate queues accordingly." [1] In Cisco IOS, PQ uses up to four queues, named high, medium, normal, and low (Figure 3) [5], and they are scheduled as shown in Figure 4. [4] The packets having the highest priority are transferred on the output port before the lower priority packets. [1] PQ uses tail-drop logic, so when a new packet arrives for a particular queue, and the queue is full, the new packet is dropped. Even if this queuing type is suitable of providing differentiated service, it also has some drawbacks, such as large continuous flow of high priority traffic into the queue, equals excessive delay, and perhaps even service starvation for lower priority packets [1].

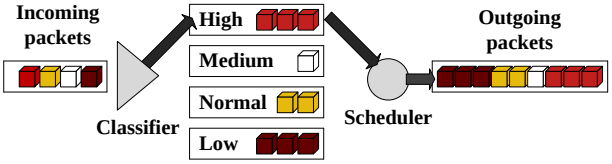

Figure 3. PQ.

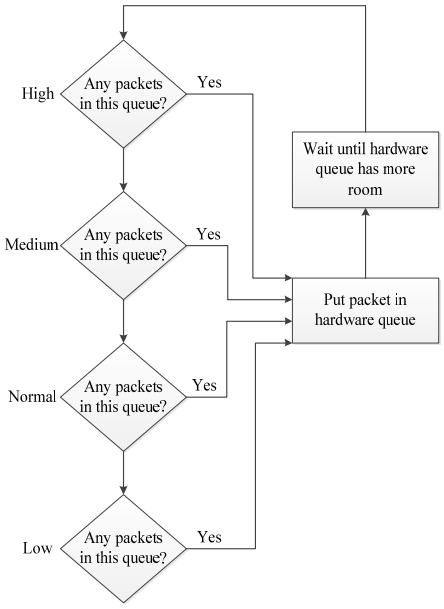

Figure 4. PQ scheduling logic [4].

IV. CUSTOM QUEUING

CQ addresses the biggest shortcoming of PQ ensuring a guaranteed minimum bandwidth to each queue, thereby queue starvation is avoided. With CQ up to 16 queues can be created by the network administrator in order to categorize traffic. (Figure 5) The queues are emptied one by one in a round-robin fashion, starting with queue 1. CQ takes packets from the queue, until the total byte count which was specified for the queue has been met or exceeded. After the queue has been serviced for the defined byte count, or when the queue does not have any more packets, CQ moves on to the next queue and repeats the process. (Figure 6)

One of the CQ queues can be setup as a default queue in order to manage traffic that is not identified specifically by the classification process. There is also one system queue which is hidden, used for important overhead traffic (routing protocol hellos, etc.). This system queue is serviced before all other queues. Cisco permits use of this queue 0, but does not recommend it. CQ uses the same classification options, and can use tail drop only for managing drops. [4]

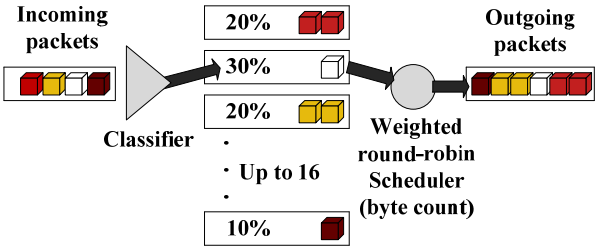

Figure 5. CQ.

The disadvantage of CQ, as compared to PQ, is the lack of a high-priority queue that is always serviced first. That is, CQ has no way to provide guaranteed low-latency service to any traffic. The CQ scheduler reserves an approximate percentage of overall link bandwidth for each queue, but instead of configuring actual percentages, CQ approximates the bandwidth percentages using a simple algorithm.

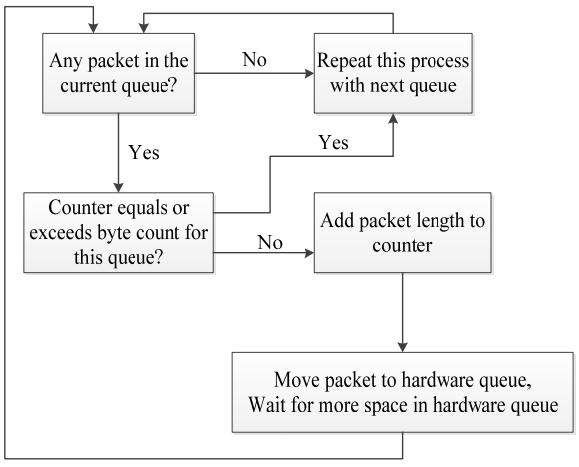

Figure 6. CQ scheduling logic for current queue [4].

V. WEIGHTED FAIR QUEUING

Processor Sharing (PS) is a type of queuing methodology having the purpose to allow fair access for each incoming flow and to prevent a bursty flow from consuming all the output bandwith. PS includes a queue for each distinct flow and packets from each flow are put into its appropriate queue. Then the system serves the queues one packet at a time using a round-robin approach.

Weighted Fair Queuing (WFQ) [6] is a subtype of Processor Sharing (PS) and it supports flows with different bandwidth requirements.[1] Weighted fair queuing differs from PQ and CQ in some significant features. Most obviously we can mention that WFQ does not allow classification options to be configured. Based on flows, WFQ classifies packets automatically, with each flow being placed into a separate queue. (Figure 7)

For the purposes of WFQ, a flow can be described as all packets with the same values for: source IP address, destination IP address, Transport layer protocol, TCP or UDP source port, TCP or UDP destination port and IP Precedence. Because WFQ puts packets of different flows in different queues, must have a greater number of queues than all of the non-flow-based queuing instruments. The WFQ scheduler uses logic that is somehow different from the logic of other queuing tools in order to be able to deal with the larger number of queues. [4]

In WFQ, there is a time stamp on each incoming packet with a finish time in addition to placing into its corresponding flow queue. In contrast to Processor Sharing, the selection of the packet to be served is now based on the time stamp on each packet. Further packets are serviced by examining their finish times. "The ones with earlier finish times are transmitted before the ones which have later finish times. It is possible for a later packet to have a finish time stamp that is smaller than an earlier packet." [1]

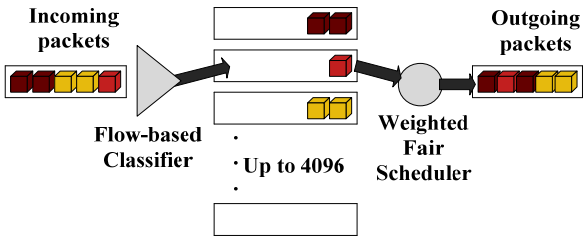

Figure 7. WFQ.

The WFQ scheduler takes the packet having the lowest finish time *(FT)* (sometimes called *sequence number*, or *SN*) when it needs to move the next packet to the hardware queue. WFQ associates to each packet an SN when the packet is added to a WFQ flow queue. The WFQ scheduler includes both the packet length and IPP when calculating the SN. The formula to calculate the SN for a packet is as below:

$$Previous\_SN + (weight * new\_packet\_length) \quad (1)$$

Where *weight* is calculated as follows:

$$weight = 32{,}384 / (IP\_Precedence + 1) \quad (2)$$

"The formula considers the length of the new packet, the weight of the flow, and the previous SN. By considering the packet length, the SN calculation results in a higher number for larger packets, and a lower number for smaller packets. By including the SN of the previous packet enqueued into that queue, the formula assigns a larger number for packets in queues that already have a larger number of packets enqueued. And by putting the weight (IPP + 1) in the denominator, packets with higher IPP values end up with lower SNs.

WFQ uses a two-step process called modified tail drop to choose when to drop packets. First, WFQ considers the absolute limit on the number of packets enqueued among all queues. This limit is called the hold-queue limit. If a new packet arrives, and the hold-queue limit has been reached, the packet is discarded. That part of the decision is based not on a single queue, but on the whole WFQ queuing system for the interface. Second, WFQ considers the length of the queue into which the newly arrived packet will be placed. Before adding a new packet to its queue, the congestive discard threshold (CDT) is checked against the actual length of that queue. If that queue is longer than CDT packets long, one packet is discarded—but maybe not the newly arrived packet. Figure 8 depicts the WFQ drop decision process.

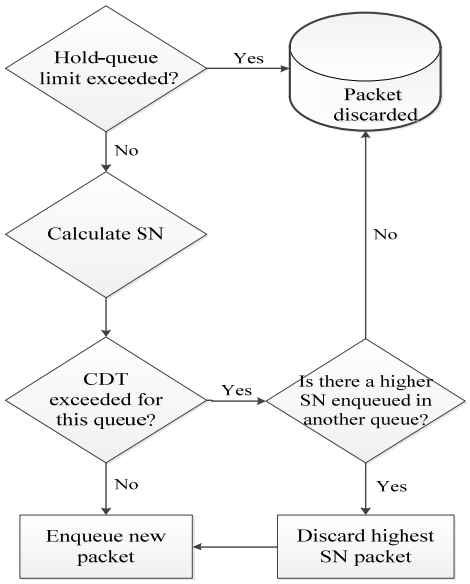

Figure 8. WFQ modified tail drop and congestive discard threshold [4].

WFQ can be configured for a maximum of 4096 queues. The allowed configurable values are powers of 2, between 16 and 4096, inclusive. IOS restricts the values because WFQ performs a hash algorithm to classify traffic, and the hash algorithm only works when the number of queues is one of these valid values. Additionally, WFQ keeps eight hidden queues for

overhead traffic generated by the router. WFQ uses a very low weight for these queues to give preference to the overhead traffic." [4].

VI. CLASS BASED WEIGHTED FAIR QUEUING

"CBWFQ [7] carries the WFQ algorithm further by allowing user defined classes, which allow greater control over traffic queuing and bandwidth allocation. CBWFQ provides the power and ease of configuration of WFQ, along with the flexibility of custom queuing. CBWFQ allows the creation of up to 64 individual classes plus a default class. (Figure 9) The number and size of the classes are based on the bandwidth. By default, the maximum bandwidth that can be allocated to user-defined classes is 75 percent of the link speed. This maximum is set so that there is still some bandwidth for Layer 2 overhead, routing traffic (BGP, EIGRP, OSPF, and others), and best-effort traffic.

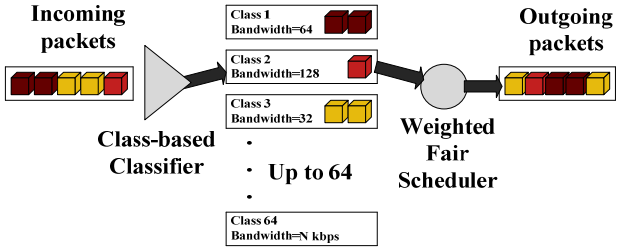

Figure 9. CBWFQ.

Each user-defined class is guaranteed a certain bandwidth, but classes that exceed that bandwidth are not necessarily dropped. Traffic in excess of the class's guaranteed bandwidth may use the "free" bandwidth on the link. "Free" is defined as the circuit bandwidth minus the portion of the guaranteed bandwidth currently being used by all user-defined classes. Within this "free" bandwidth, the packets are considered by fair queuing along with other packets, their weight being based on the proportion of the total bandwidth that was guaranteed to the class. [8]

All packets not falling into one of the defined classes are considered part of the default class. The default class can be configured to have a set bandwidth like other user-defined classes, or configured to use WFQ in the remaining bandwidth and treated as best effort. When the fair queuing buffers overflow, packets are dropped with tail drop unless WRED (Weighted Random Early Detection) [9] has been configured for the class's policy. In the latter case, packets are dropped randomly before buffers totally run out in order to signal the sender to throttle back the transmission speed." [4]

VII. LOW LATENCY QUEUING

Neither WFQ nor CBWFQ can provide guaranteed bandwidth and low-delay guarantee to selected applications such as VoIP. "That is because those queuing models have no priority queue. Certain applications such as VoIP have a small end-to-end delay budget and little tolerance to jitter. LLQ [10] includes a strict-priority queue that is given priority over other queues, which makes it ideal for delay and jitter-sensitive applications. Unlike the plain old PQ, whereby the higher-priority queues might not give a chance to the lower-priority queues and effectively starve them, the LLQ strict-priority queue is policed. This means that the LLQ strict-priority queue is a priority queue with a minimum bandwidth guarantee, but at the time of congestion, it cannot transmit more data than its bandwidth permits. If more traffic arrives than the strict-priority queue can transmit, it is dropped. Hence, at times of congestion, other queues do not starve, and get their share of the interface bandwidth to transmit their traffic. Figure 10 shows an LLQ.

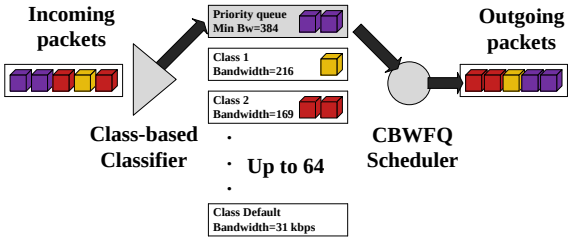

Figure 10. LLQ.

As we can see, LLQ is effectively a CBWFQ with one strict-priority queues added. It is possible to have more than one strict priority queue. This is usually done so that the traffic assigned to the two queues – voice [11], [12] and video traffic, for example - can be separately policed. However, after policing is applied, the traffic from the two classes is not separated. It is sent to the hardware queue based on its arrival order (FIFO). As long as the traffic that is assigned to the strict-priority class does not exceed its bandwidth limit and is not policed and dropped, it gets through the LLQ with minimal delay. This is the benefit of LLQ over CBWFQ." [3] The LLQ scheduler logic is shown in Figure 11.

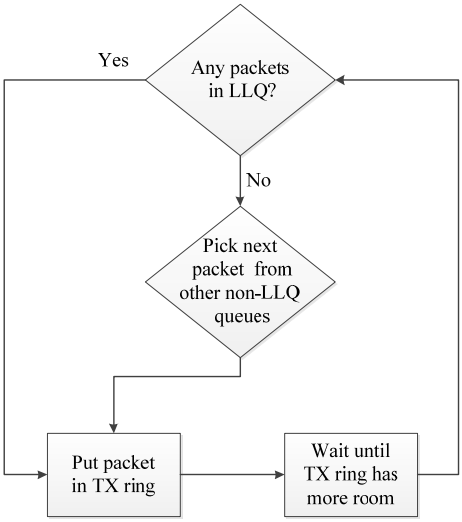

Figure 11. LLQ scheduler logic [4].

VII. CONCLUSIONS

Table 1 summarizes some of the key points regarding the IOS queuing tools covered in this paper.

TABLE 1. Queuing protocol comparison [4].

Feature	**FIFO**	**PQ**	**CQ**	**WFQ**	**CBWFQ**	**LLQ**
Includes a strict-priority queue		Yes				Yes
Polices priority queues to prevent starvation						Yes
Reserves bandwidth per queue			Yes		Yes	Yes
Includes robust set of classification fields					Yes	Yes
Classifies based on flows				Yes	Yes[2]	Yes[2]
Maximum number of queues	1	4	16[1]	4096	64	64

[1]Also includes a system queue that is unavailable for customer use.
[2]WFQ can be used in the class-default queue, or in all CBWFQ queues in 7500 series routers.

REFERENCES

[1] T. Svensson, A. Popescu, "*Development of laboratory exercices based on OPNET Modeler*", Master thesis, Blekinge Institute of Technology, Department of Telecommunications and Signal Processing, 2003

[2] QOS, "*Implementing Cisco Quality of Service*", Student Guide, Volume 2, Version 2.2, © 2006 Cisco Systems Inc.

[3] A. S. Ranjbar, "*CCNP ONT Official Exam Certification Guide*", Cisco Press, 2007

[4] W. Odom, J. Geier, N. Mehta, "*CCIE Routing and Switching Official Exam Certification Guide*", Second Edition, Cisco Press, 2006

[5] "*Cisco IOS Quality of Service Solutions Configuration Guide*", Release 12.4T, © 2008 Cisco Systems Inc.

[6] M. Barreiros, P. Lundqvist, *"QOS-Enabled Networks - Tools and Foundations"*, A John Wiley & Sons, Ltd., Publication, 2011

[7] W. Odom, R. Healy, D. Donohue, *"CCIE Routing and Switching Certification Guide"*, 4th Edition, Cisco Press, 2010

[8] B. Durand, J. Sommerville, M.Buchmann, R. Fuller, *"Administering Cisco QoS in IP Networks"*, Syngress Publishing, Inc., 2001

[9] R. S. Benn, S.C. Kronenberg, E. Rozell, *"Configuring Cisco AVVID – Architecture for Voice, Video, and Integrated Data"*, Syngress Publishing, Inc., 2001

[10] K. Wallance, *"Authorized Self-Study Guide - Cisco Voice over IP (CVOICE)"*, 3rd Edition, Cisco Press, 2009

[11] P. K. Verma, L. Wang, *"Voice over IP Networks - Quality of Service, Pricing and Security"*, Lecture Notes in Electrical Engineering, Vol. 71, Springer, 2011

[12] S. Kashihara, *"VoIP Technologies"*, INTECH Open Access Publisher, 2011

Online Software Module for Measurement of Audio-Frequency Controlled Heat-Storage Power

UNHAUZER Attila, VÁRADINÉ SZARKA Angéla

University of Miskolc, Hungary,
Department of Electrical and Electronic Engineering, Faculty of Mechanical Engineering and Informatics,
Miskolc, H-3515 Miskolc-Egyetemváros, Hungary,
E-mail: elkua@uni-miskolc.hu, elkvsza@uni-miskolc.hu

***Abstract* –Efficient and direct power control stands in the centre of electrical engineering research. One possibility for influencing power consumption is the control of heating units. The direct control of heat-storage power is realized through centralized ripple control system, which is a simple, traditional and efficient solution in the energy management from the power suppliers' point of view. The audio-frequency and radio-frequency controlled systems use only one-way signal transmitting, contrary to SMART systems which ensure the control and measurement as well while sending feedback information from the customer-units. The development of SMART systems on customer side is very expensive; in addition, research in this topic is in the phase of testing. Consequently, it raises the question: how the centralized ripple control system can "receive" online information about heating power. This paper summarizes the results of our research running in Northern Hungary aimed at developing an online module for measurement of audio-frequency controlled heat-storage units.**

***Keywords:* centralized ripple control, stochastic power fluctuation, PC-based measurement, power control on electrical network, demand side management.**

I. INTRODUCTION

The requirements of power supply are regulated by *Parliamentary Act LXXXVI of 2007 on Hungarian Electric Energy* [1]. The power companies have to observe various requirements due to strong competition on the market. They are interested in high quality electrical energy and obviously, they would like to maintain favorable economic conditions at the same time.

The type, power and number of machines used by consumers stochastically vary on the electrical network, which can easily be identified on daily energy consumption throughout the day.

Energy suppliers buy electrical energy on the international market; however, they cannot accurately calculate the level of daily energy consumption. It is a fact that suppliers have to inform the *Hungarian Transmission System Operator Company Ltd. (MAVIR)* about the power consumption changes on the network on a 15 minute-basis, during a calendar day. The daily energy consumption can be defined with the help of a special *power load curve* which shows the consumption data throughout a day. In the event of wrong energy approximation, the power supplier has to pay penalty (this economic burden is borne by the consumers), therefore, the accurate estimation of power consumption is of high importance [2].

Through *audio-frequency (AF) control (AFC)*, the different suppliers (e.g.: *ÉMÁSZ: North Hungarian Electricity Supply plc.; ELMŰ: Budapest Electricity Works*) can influence maximum 10 % of the whole energy consumption on the electrical network. The load profile can be directly handled through controlling the heating units (e.g.: boilers, heating stoves) by the supplier in order to keep the required power-scheduling.

The key element of this project was to develop a special *online software module* in order to determine the heating power more accurately by the suppliers than before; since previously, they used only empirical estimations for defining the power of heat-storage units. The present research work consists of five different phases which are outlined in the followings:

- Studying the *audio-frequency controlled system (AFCS)* in Hungary [3];
- Measurements and data acquisitions at 35 different transformer stations of *ÉMÁSZ;*
- Development of new and efficient analyzing methods; development of analyzing software working offline mode [4-7]; and
- Development of an online module for complex analyzing software to enable the computation of heat-storage load continuously.

II. ARCHITECTURE AND OPERATION OF AFCS

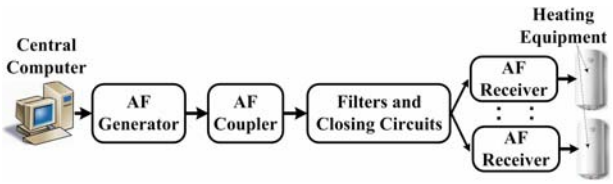

Figure 1. Schematic structure of AFCS.

AF generators (Fig. 1) start AF messages at each transformer station. First, generated AF messages have to be coupled and filtered before the signal will be modulated to the high voltage network. These highly-efficient generators (100 kW-1 MW) create 3-phase AF signal to the network (depending on the network capacity). ÉMÁSZ uses 183.33 Hz for centralized ripple control switching on the electrical network. Other suppliers use different frequencies for AFC, e.g.: ELMŰ uses 213.66 Hz. The amplitude of generated AF signal is between 1% and 2% the mains voltage.

AF coupler modules (Fig. 1) are the most crucial elements of the AFCS. The previously generated AF signal will be coupled to the high-voltage network (50 Hz) in serial and parallel mode (ÉMÁSZ: 10 kV, 20 kV; ELMŰ: 120 kV). The *filters and closing circuit modules* decrease both damping and harmonic distortion of the AF signal.

AF receiver modules (Fig. 1)are installed on the customer side and connected to the electrical network (230 V). Receivers can switch on or off the heating equipment by relays. Electrical circuits of receivers are programmed before installation. AF receivers can be started by individual AF messages. ELMŰ and ÉMÁSZ use different AFCS (ÉMÁSZ: Pulsadis System; ELMŰ: RiconticB System with Versacom Protocol). Both systems apply different structures of AF messages, with different lengths of pulse.

The *central computer module* (Fig. 1) controls AFCS. AF generator modules at transformer stations (in different geographical regions) send AF messages (a series of pulses, about 60 seconds long) according to special scheduling. AF messages are sent in well-defined time intervals (usually in every 15 minutes) and are received by customer equipment, thus boiler heating starts or stops. These messages are synchronized and sent to the electrical network in order to provide accurate switching shifting.

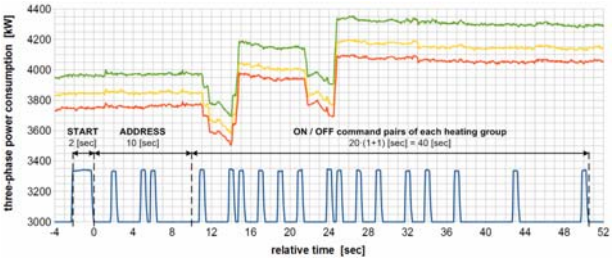

Figure 2. Structure of generated AF messages.

One message (in the ÉMÁSZ system) consists of three elements (*Fig. 2*): start pulse (start pulse lasts about 4 sec long); followed by addressing, which includes 10 pulses (the address determines the type of customer equipment, and each pulse lasts 1 sec with a 50% duty cycle); and finally 20 command pairs to switch the 10 boiler groups on or off. The boiler clusters are controlled by individual command pairs inside an AF message. The structure of these commands is very simple, the first pulse switches the first group on, the second pulse switches the first group off, the third pulse switches the second group on, the fourth switches it off, and so on.

Various AF messages, differing in delay and structure, are used for controlling customers' equipment directly by power suppliers. Consequently, numerous variations of the sequence and length of AF pulses occur to influence the receivers on customer side.

III. QUERIES OF AFCS

Obviously, numerous theoretical-, and practical questions arise during the control of heating equipment (during the AFC).

- After sending of *AF* messages, boilers do not turn on in certain cases; therefore, difficulties can be identified during the prediction of (stochastic) power changes on the electrical network. This phenomenon cannot be handled by the suppliers, since various reasons can exist in the background (e.g. fully charged boilers, unrecognizable messages on costumers' side).
- Throughout a day, water-consumption continuously and stochastically changes, moreover, the energy needed for boiler-heating changes in the same way as well.

Suppliers are not aware of up-to-date and accurate power data about *"built-in controlled power"* (*heating load*) and they do not keep central registration about the number of heating equipment. In addition, they know only estimated data about the boiler-power on customers' side. Consequently, even if the suppliers kept accurate registration about boilers on the network, they would not be able to estimate the instantaneous heating power correctly, since it is influenced by various external circumstances as well. Furthermore, suppliers do not know the level of heating power changes, which occur due to switching on and off boilers in the regions controlled by transformer stations.

According to the *one-way signal controlled AFCS*, the suppliers have only one possibility to obtain useful information about their own electrical network: measurements have to be carried out in the different AF-regions in order to examine the power load of heat-storage units. Suppliers do not have any feedback about heating units, consequrently, we developed a special online software module using new methods for the analysis of the network. Thus, suppliers are able to obtain useful and precise information about their own electrical network. These developed methods can be easily adjusted to modern radio-frequency controlled systems as well.

IV. ACCOMPLISHED MEASUREMENTS AND NEW DEVELOPED ANALYZING METHODS

First of all, on 35 different transformer stations of ÉMÁSZ (altogether 45 transformers) measurements were carried out. These measurements characterized all suppliers of Hungary well; thus, a universal online module could be developed. In order to be able to

recognize AF messages and to determine the 3-phase power on the network, 3-phase voltages and currents were generally measured using 3200 Hz sampling frequency during the PC–based data acquisition (Fig. 3). These measurements were performed for 150 days and more than 15,000 AF messages were analyzed on the electrical network.

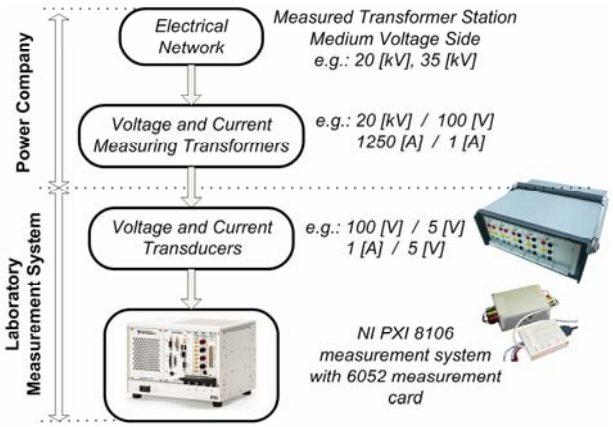

Figure 3. Modular structure of the applied measurement system.

As a result of our measurements, more than 600 GB data were acquired; therefore, the data analysis had to be automated for easy evaluation of files. Consequently, brand new methods had to be developed to analyze the measured data. During the development process, three critical problems had to be solved and ultimate complex analyzing software was created for studying the operation of AFC. The diagnostics software helps ÉMÁSZ to study its electrical network in offline mode.

First of all, the *recognition of AF pulses* was a critical problem, therefore, we had to develop a reliable detecting method. During this research, a simple frequency analysis method was applied to recognise AF messages. We determined the *optimal size of FFT–time window* which was necessary for the reliable detection of AF pulses.

Secondly, the *recognizing method* was refined, as it was automated, being developed for *detecting AF messages* (pulses) and for the *computation of 3-phase power*. The developed methods save the calculated data in a structured matrix-database, thus making the further analysis more efficient. As a conclusion, AF pulses could be identified with 99.94% reliability (3 periods size FFT-window) by the automated detection method.

Finally, after the recognizing process of AF pulses, the most important step was to determine the power fluctuations caused by AFC. Two significant delay characteristics were identified during the processes of boilers' switching on or off. Based on our experiment, these parameters primarily influence the computation of power fluctuation (heat-storage power). The two critical parameters (Fig. 4), which can influence the power calculation, are:

- the *reaction time of the customer receivers* (the time parameter between the moments of the AFC command and the equipment's real switching on or off) and
- the *operational delay of significant heating power fluctuation* after the start of equipment.

In most cases the perceptible power changes tend to start about 0.1-0.3 sec after the start of AF pulse and stop 0.8-1.2 sec after the pulse ends. This means that the most important setting during an automated estimation process is the size of the monitoring window for power changes, to simplify the perception of 3-phase power fluctuation on the electrical network (Fig. 5).

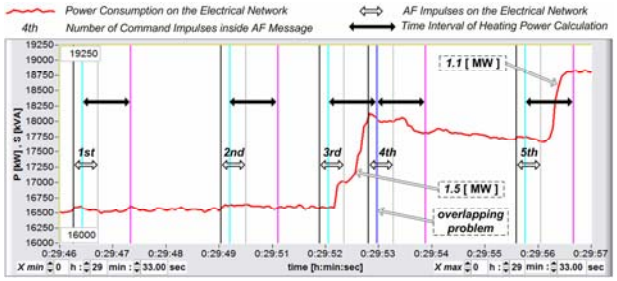

Figure 4. Introducing the computation method of heat-storage power.

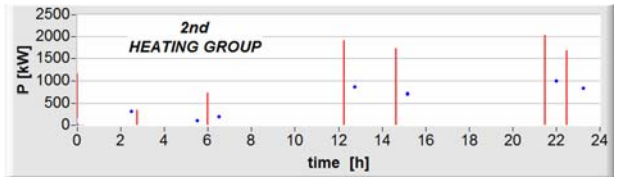

Figure 5. Displaying switched on (*vertcal lines*) and off (*solid points*) heating power changes by 2nd boiler group (transformer station of *Károlyfalva*).

V. ONLINE SOFTWARE MODULE DEVELOPMENT FOR ESTIMATION OF HEAT-STORAGE POWER

An online software module was developed due to the requests of ÉMÁSZ for detecting AF messages, for estimating the heat-storage power and for saving the results [8].

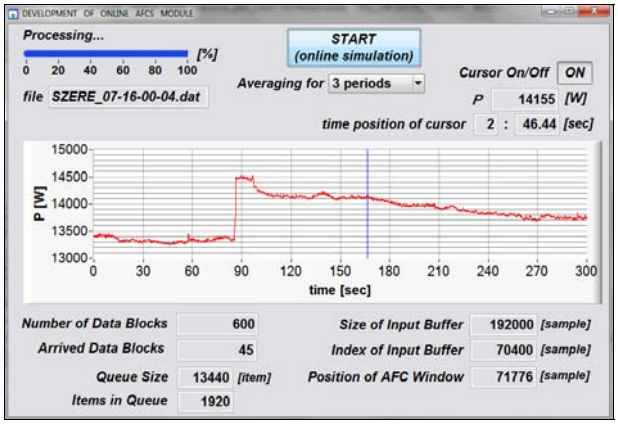

Figure 6. User interface of the developed online simulation software displaying the real-input data.

Thus, we had to create a special simulation application for studying the operation of online (final) software (Fig. 6). The developed application simulates a

general data acquiring process of the electrical network using simple data files from one of the earlier measurements. Furthermore, the software uses real acquired data which can properly simulate the operational phases of real measurements. The developed simulation utilizes a modern *multi-core data processing method with advanced thread communication* [9-10]. Three different threads are applied for communication (Fig. 7).

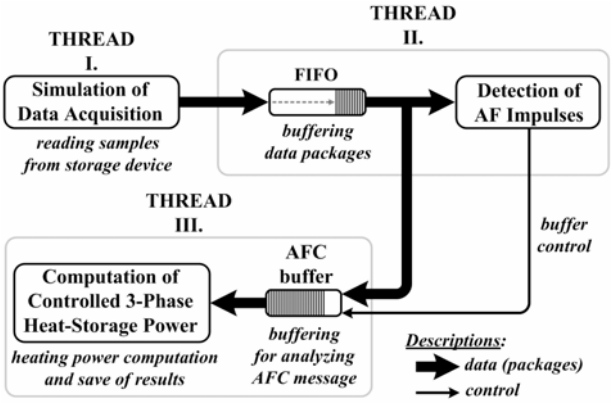

Figure 7. Schematic structure of threads' communication applying online simulation.

Every thread works parallel and is synchronized with the others. After the *data acquisition thread* (*Thread I*) begins to read the measured data, sends it to *FIFO* (*first in first out buffer*); then *Thread II* can recognize the AF pulses. In case of detecting a start AF pulse, *Thread II* permits the sending of data into the *AFC buffer,* facilitating the computation and saving of 3-phase heating power. Although we had tested recognition and computation methods before, several synchronization issues had to be solved during the development of the online simulation, e.g.: the size of *FIFO* or the synchronization between *Thread II* and *Thread III.*

After the tested simulation, the online module had to be built in complex data acquisition software. For the purpose of the future easier integration of the online module, well-defined module interfaces had to be created simplifying the further communication among the modules.

VII. CONCLUSIONS

This paper summarizes our present research achievements. Particular attention was paid to introduce our earlier accomplished measurements, along with different development and estimation processes; furthermore, the creation of an online module was briefly outlined as well.

Before the launch of our research, ÉMÁSZ did not receive any feedback from AFC units and it could only estimate the heat-storage power on its network. Thus, we put emphasis on the reliable determination of power changes. With the help of computed power changes, exact *heat-storage power profiles* could easily be approximated, therefore, ÉMÁSZ can optimize its electrical network.

Another main advantage of our developed online module can be summarized as follows: currently, *Hungarian power suppliers* can control heating equipment only through *audio-frequency* and *radio-frequency controlled system*. Our developed online module helps suppliers to identify the heating power on their electrical network. Although this method was developed for AFCS, it can easily be converted to RFCS as well. It is of crucial importance that only centralized ripple control solutions are going to spread in Hungary in the next decade, while SMART systems may not spread rapidly in the near future.

ACKNOWLEDGEMENTS

The described work was carried out as part of the *TÁMOP-4.2.1.B-10/2/KONV-2010-0001* project in the framework of the *New Hungarian Development Plan*. The realization of this project is supported by the *European Union*, co-Financed by the *European Social Fund*.

REFERENCES

[1] *Requirements of power supply are regulated by Act LXXXVI of 2007 on Hungarian Electric Energy* (Promulgated on 2 July 2007): http://www.eh.gov.hu

[2] P. Honey: "*Electric water-heating development with particular reference to domestic use*", IEE Electrical Engineers Proceedings, vol. 112(10), pp. 1951-1961, 2010.

[3] A. Neuberg: "*Ripple control in the Czech Republic and demand side management*", Electricity Distribution - Part 1 (CIRED 2009 Proc), 20th International Conference and Exhibition.

[4] B. Rautenbach, I. E. Lane: "*The multi-objective controller: a novel approach to domestic hot water load control*", IEEE Transactrion of Power Systems, vol. 11(4), pp. 1832-1837, 1996.

[5] H. Salehfar, A. Wehbe: "*Direct control of residential water heater loads to reduce power system distribution losse*s", IEEE Proceedings Power Engineering Society Winter Meeting, vol. 3, pp. 1455-1460, 2001.

[6] I. Lane, N. Beute: "*A model of the domestic hot water load*", IEEE Transaction on Power Systems, vol. 11(4), pp. 1850-1855, 1996.

[7] E. F. Lemmer, G. J. Delport: "*The influence of a variable volume water heater on the domestic load profile*", IEEE Transaction on Energy Conversion, vol. 14(4), pp. 1558-1563, 1999.

[8] M. H. Nehrir, B. J. LaMeres, V. Gerez: "*A customer-interactive electric water heater demand-side management strategy using fuzzy logic*", IEEE Proceedings of Power Engineering Society 1999 Winter Meeting, vol. 1, pp. 433-436, 1999.

[9] T. C. Chieu, Zeng Liangzhao: "*Real-time performance monitoring for an enterprise information management system*", IEEE International Conference on e-Business Engineering (ICEBE '08 Proc), pp.:429-434, 22-24 October 2008.

[10] M. U. Kobe, A. C. Tsoi: "*Modelling of domestic hot-water heater load from online operating records and some applications*", IEE Proceedings of Generation, Transmission and Distribution, Auckland, New Zealand, vol. 133(6), pp. 336-345, 11 November 2008.

Design and Dynamic Analysis of Six Legged Walking Robot

YILDIRIM Şahin, ARSLAN Erdem

Erciyes University, Turkey,
Department of Mechatronics Engineering, Engineering Faculty,
38039 Kayseri, Turkey, E-Mail: sahiny@erciyes.edu.tr, erdemarslan@erciyes.edu.tr

Abstract *– This paper presents an investigation for designing and dynamic analysis of six-legged walking robot. Firstly, the robot is designed by software for predicting desired force on inertial parameters. Secondly; dynamical walking conditions are tested with open dynamics engine software. By applying this acquired incitation components to the dynamic model on simulation software, it has been performed a comparison between the reference trajectory used on calculations that made before dynamic analysis and the trajectory ingenerated from the model after dynamic analysis.*

Keywords: *legged robot design; dynamic analysis; dynamic simulation; walking robot.*

I. INTRODUCTION

With developing technological opportunities the numbers of fields mobile robots can be used on increases. From unmanned air vehicles to the robots that send to extraterrestrial planets, mobile robots could find an opportunity to commonly using. By using their peripheral equipments mobile robots perform their function properly even changing environmental conditions.

Legged robots have important role among mobile robots on account of achieving their serve in quite difficult field conditions. The US army estimates the wheel can only reach %50 of the places on earth [1]. In contrast to wheeled systems the ability to practice their job even in mountainous terrain make possible to use walking robots in military operations like mine clearance, strike and defense. Mostly by use of biomimetic designs in legged robots, required movements that need to be done supplied rapidly and correctly.

Due to high technology on robotic systems there is still some problems on walking conditions of legged robots. The first of these the required dynamic calculations for legged mobile robots movement control are vitally complex and computing demand for solving these nonlinear equations is overmuch. In recent years with raising processor speeds it has been made to the solution partially. But for the dynamic calculations of the legged mobile robots has higher degrees of freedom (DOF) overmuch burden on process required. Another basic problem is no suitable power and incitation systems could be performed yet to the legged robots need too much power demand in proportion to wheeled mobile robots.

Biomimetic designs are widely using on legged mobile robots. Researchers study on animals from cockroach [2] to small animals [3] for achieving well structures that have ability to make sufficient movements. On the other hand some researchers choose using technological opportunities like artificial tendons [4] or springs and dampers [5] to design legged robots. Zielinska and Heng [6] select the way of employ the experimental results on design. For this reason they create a system that gives possibility to try all the configurations (insect, mammalian or reptile) and part dimensions.

To perform dynamic analyses in simply way countless methods tried to develop till now. Especially systems hard for dynamic modeling miscellaneous approaching methods would be tried on. Among the approaching methods some studies is existed that dynamic action not to be taken into consideration because of the lower dynamical effects [7-9]. In these studies by using methods known as gait analysis it has been aimed to move the trunk body in a fast and stable way.

II. DESIGN OF ROBOT

In mobile robot design, it is significant presenting system design that has ability to perform its movement effectively. In this context various parameters like part number, weight, leg design needs to be chosen canorously. Determining higher leg numbers gives to system more stable and steady movement ability, in spite of that it also raises the weight. In the event of weight rising, the required incitation power of system elements drives the body increases proportionally. The rising on power requirement, could causes problems on legged robots that has feature like mobility and flexibility.

In walking robot design six legged configuration that seen on insects preferred in the name of reducing probability of unbalance situations. In case of malfunction of any leg, symmetrically approach is followed on positioning legs upon the body. Joint number required for each leg is chosen three in order to performing movement in three dimensional space.

The most important parameter that needs to be defined on designing level is the ratio of tibia and femur parts lengths. Fundamentally there are two approaches exists on dimensioning this ratio for robot design. The first of these is researching the creatures on nature to obtaining a correlation between a leg speed and the force that the leg could supply. On the second approach lowly power required configurations chosen in the light of data that acquired from experimental studies [6]. The mechanism of researchers tried various configurations on is given in Fig.1.

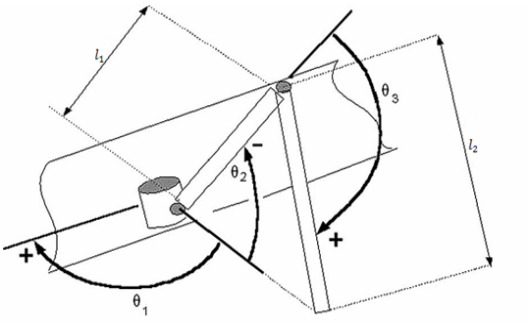

Figure 1. Schematic representation of leg system [6].

To understanding the first method, limbs can be seen as levers. Long limbs give a high velocity ratio [but less output force (leverage)] and so favor running; short limbs have more output force (greater leverage) but a lower velocity ratio, so are suitable for activities such as digging [25]. Because of the necessity to force is more importer than the velocity on legged mobile robots the dimension of tibia and femur parts chosen closely.

On the second method experimental results could be using to define tibia and femur ratio. Researchers aimed at lower incitation power for motors using in joints and for this aim they tested seven different configurations in insect, reptile and mammalian type. The results of Zielinska and Heng's [6] research are given in Fig.2.

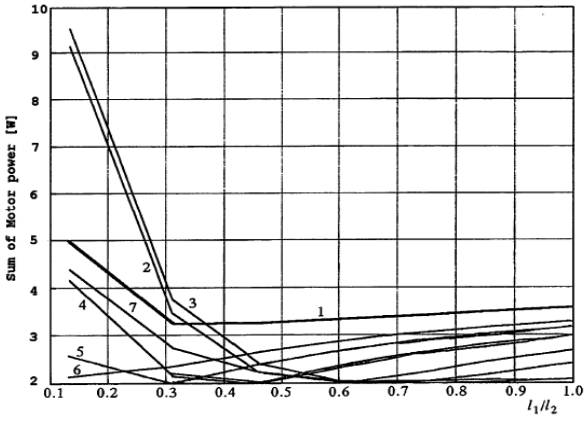

Figure 2. Leg results of Zielinska and Heng's research [6].

Another fascinated method for defining tibia-femur ratio is using golden ratio. From dimensions in human hand to snail shape, a lot of parameters could explain by using of golden ratio in creatures lives on nature. Theoretical and experimental results in the research up to this point take in consideration on determining tibia-femur ratio. In the course of design this ratio is chosen 0.61 that is golden ratio divided into one. By this means both part dimension chosen closely for raising the force supplement like in digging, and the experimental results from Zielinska and Heng presented taken into account. After these definitions the design completed in CAD program. The motors and batteries included to design to embody the dynamic effects of this elements during the simulation. The three dimensional drawing of this design is outlined in Fig. 3.

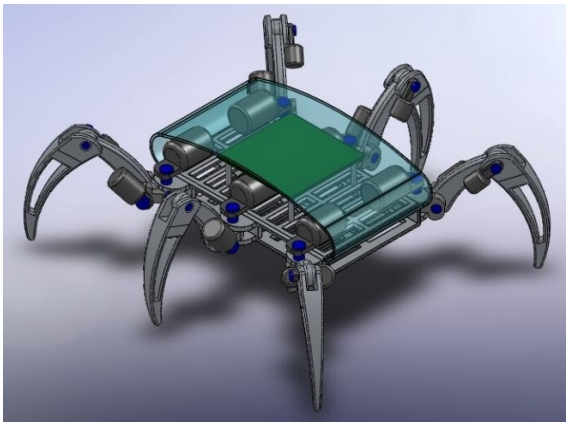

Figure 3. Designed six legged walking robot.

Aluminum alloy (1060) is employed on the legged robot design in the name of raising strength of mechanical system by using of lesser weight. After the design the legged mobile robot; consist of 19 parts and 18 degrees of freedom (DOF); has 6.87 dm³ volumes, total weight of this robot is 18.5 kg, the mass of the body is 9.2 kg and the mass for each leg is 1.55 kg.

III. DYNAMICS OF ROBOT

Calculation of the incitation forces required for movement according to systems dynamic parameters is the basis of robot dynamics. Additionally to stability condition in statically case, it is need to be consider the dynamic effects arise from feature of mass and shape. Therefore before starting to dynamic analysis it must be defined dynamic parameters take its source from the shape and structural characteristics of the parts using in legged mobile robot. Basically, the dynamic model of a robot manipulator finds mathematical relationships among:

- Robot location and its derivatives, velocity and acceleration.
- Forces and torques applied at the robot joints or end-effectors.
- Dimensional parameters of the robot manipulator, such as link length, mass and inertia.

On walking mobile robots legs connected to each other by using both with body and ground as a closed kinematical chain. Therefore it is need to be determined contact forces between the leg and surface beside the

other parameters. As a result of the dynamic analysis starting from the D'Alambert's principle (dynamic stability) in general, dynamic equation that dynamic effects could determine tried to be deriving in standard form. It can be expressed in the following equation,

$$\tau_e - J^T F = D(q)\ddot{q} + H(q,\dot{q}) + C(q) \quad (1)$$

On the equation q term stands for system variables in generalized coordinate system. D function is a square matrix that indicates dynamic effects of the mass in equation. H function stands for Coriolis and Centrifugal effects on equation.

C function represents the gravitational effects and τ term stands for the forces applied to system. F term exists for external forces and J matrix is situated for transformation of these forces known as Jacobian matrix.

On the other hand there are some simplifications can be used like higher reduction rates in joints for ignoring Coriolis and Centrifugal effects. But this method make calculations easy cannot make a fully solution because of the friction and elasticity in joints.

IV. DYNAMICS OF ROBOT

Newton-Euler algorithm is used for the dynamic analysis of legged mobile robot. Among the other methods this algorithm is using because of the simplicity and rapidity. As mentioned at before this algorithm consists of two phases. At the first phase of algorithm positions, velocities and accelerations computed from reference trajectory.

After determining these parameters incitation elements calculated using D'Alambert's law and the inertial forces during the motion. For achieving speed on dynamic calculations, the first phase knowing as inverse dynamics calculated to fixed step sized ODE.

At the first phase of this algorithm inverse kinematics need to be solved in the first place. Inverse kinematics problem is not always be solvable as closed form equation, even in industrial robots that fixed a point. It is hard to be overcome this problem in legged mobile robot that has eighteen degrees of freedom with non-holonomic constraints.

Also jacobian matrix define the velocity propagation is not easy to derive. In this algorithm pre-runned inverse dynamics model is using on determining velocity propagation and contact forces in case of analytical methods. From the trajectory defined by the help of ODE; required data's like joint-part positions, velocity propagation and contact forces which will occur between the feet and surfaces; have been obtained at the end of the inverse dynamics.

After inverse dynamics calculations in ODE; dynamic stability law written in the method defined as guided running algorithm and the torque elements needs to be applied for achieving the reference trajectory is obtained. Afterwards these torque values applied to the robot dynamics (forward) modeled in ODE and by application of these torque values a motion is observed similar to reference trajectory. The block diagram of guided running algorithm is outlined in Fig. 4.

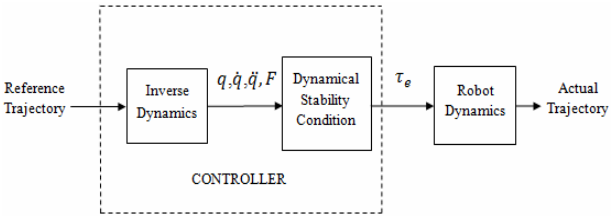

Figure 4. Block diagram of guided running algorithm.

On behalf of predetermining five-step reference walking trajectory for thirty second simulation; position, velocity and acceleration distributions charted as a time dependent function for all parts. Additionally to these distributions ODE using for determining contact forces ingenerated at the end of feet by using sensors. By reason of all the parts and its configurations is the same, formulation presented just on a one leg.

After the calculation of eighteen torque values each required for all degrees of freedoms data's recorded as a function of time. The incitation and inertial forces using in Newton-Euler algorithm on leg configuration is outlined in Fig.5.

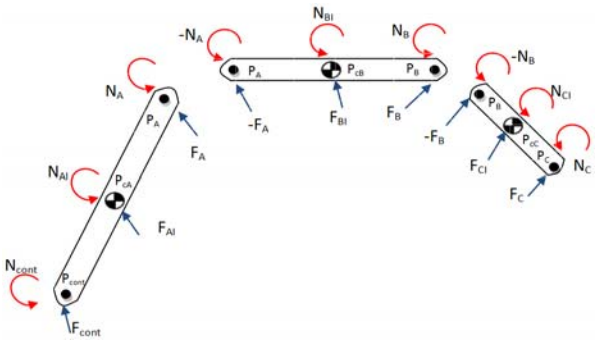

Figure 5. The incitation and inertial forces on leg configuration.

After acquiring the position, velocity and acceleration data's as function of time, by writing a stability law for each part bonding forces obtained. For that purpose equation (2) and (3) applied to parts in turn from outside to inside. From starting at the end of feet contact forces accepted as an external force and inertial forces for each part concluded to equation respectively.

$$N_i = N_{A_i} + {}_{i+1}^{i}R^{i}.N_{i+1} + \tilde{I}_{A_i}^{(i)}\alpha_{A_i} \times F_{A_i} + \\ + \tilde{I}_{i+1}^{(i)}\alpha_{i+1} \times {}_{i+1}^{i}R^{i}.F_{i+1} \quad (2)$$

$$F_i = {}_{i+1}^{i}R^{i}.F_{i+1} + F_{A_i} \quad (3)$$

On the equations N term stands for torques and F term stands for linear forces at the point. R term is a matrix used for rotation transformation between the frames (direction cosines) and P term is a vector for frames displacements. After obtaining the bonding forces constraint matrix (Z) is needed for computing incitation torques required for performing desired

motion from In pursuit of multiplying (Eq. 4.) constraint matrix and bonding torques (N) at the joints; the incitation torque(τ) required for desired motion derived.

$$\tau_t = N_t^T . \mathbb{I}(t) \ddot{\Xi}_t \quad (4)$$

Afterwards by applying; the derived torque values obtained from dynamic stability law; to dynamic model (forward dynamics) in ODE; the open-loop results of this algorithm achieved. The dynamic model of the robot in ODE is shown in Fig. 6. To understanding the quality of proposed guided running algorithm the reference trajectory and the ingenerated trajectory at the bodies center of the mass (CoM) outlined in Fig. 7.

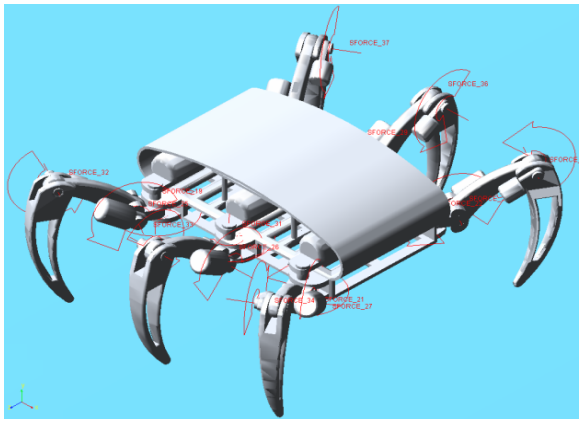

Figure 6. The dynamic model of the robot in ODE.

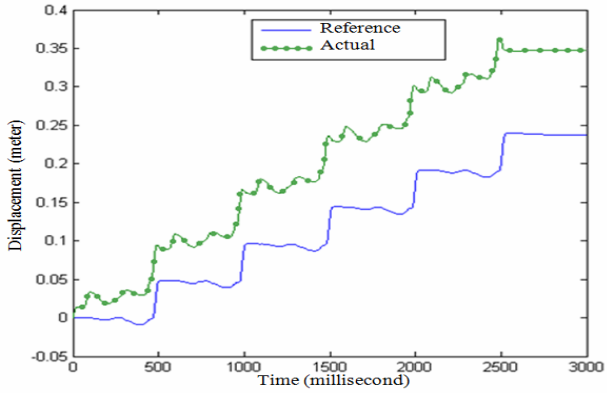

Figure 7. Trajectory variations of robot's center of mass.

As can be seen from the graph shown in Fig. 7. a motion similar to reference trajectory performed. As a result with this proposed guided running algorithm; inverse dynamics could place a burden on processor speed solved by using of generalized ODE that use fixed step sized solver. Besides it is also possible to use this algorithm (open-loop structure shown in Fig. 4.) with feedback control systems by adjusting the fixed step size on solver. Therefore it is thought to be feasible proposed computation method on application of real-time legged mobile robot. acceleration distribution in inverse dynamics calculations had been solved.

V. CONCLUSION AND DISCUSSION

This paper has presented to design of six-legged walking robot for pre-described leg trajectories. On the level of robot design parameters tried to be adjust like part dimension by taking into consideration the trade-off problem between the speed and force of the leg. By the parameters that obtained from the design a model tried to be developing for solving the dynamic analysis problem named proposed guided running algorithm.

With using the fixed step sized solvers the propagation that could occur on inverse dynamics using for the first phase of Newton-Euler Algorithm endeavor to prevented. Performing of this dynamic analysis in a swiftly way, make the proposed algorithm possible on application robot works in real-time conditions. As a result of thirty second dynamic simulation position, velocity and acceleration data's obtained and after application of these data's to the proposed guided running algorithm in open loop way, similarly movement to the reference trajectory resulted with some undesired drift. It is expected that the drifting error problem in open loop system will be solved by using feedback control systems like PID, NN (Neural networks) or ANFIS (Adaptive neuro-fuzzy).

REFERENCES

[1] Barreto, J., Trigo, A., Menezes P., Dias, J., Almeida, A.T., 1998. FED-the free body diagram method kinematic and dynamic modeling of a six legged robot. Advanced Motion Control, AMC '98-Coimbra International Workshop: 423–428.

[2] Delcomyn, F., Nelson, M.E., 2000. Architectures for a biomimetic hexapod robot. Robotics and Autonomous Systems, 30: 5-15.

[3] Scarfogliero, U., Cesare, S., Dario, P., 2009. The use of compliant joints and elastic energy storage in bio-inspired legged robots. Mechanism and Machine Theory, 44: 580-590.

[4] Ghorbani, R., Wu, Q., 2009. Adjustable stiffness artificial tendons: Conceptual design and energetics study in bipedal walking robots. Mechanism and Machine Theory, 44: 140-161.

[5] Chatzakos, P., Papadopoulos, E., 2009. Bio-inspired design of electrically-driven bounding quadrupeds via parametric analysis. Mechanism and Machine Theory, 44: 559-579.

[6] Zielinska, T., Heng, J., 2003. Mechanical design of multifunctional quadruped. Mechanism and Machine Theory, 38: 463-478.

[7] Klaassen, B., Linneman, R., Spenneberg D., Kirchner, F., 2002. Biomimetic walking robot SCORPION: Control and modeling. Robotics and Autonomous Systems, 41: 69-76.

[8] Erden, M.S., Leblebicioğlu, K., 2007. Free gait generation with reinforcement learning for a six-legged robot. Robotics and Autonomous Systems, 56: 199-212.

[9] Pratihar, D.K., Deb, K., Ghosh, A., 2002. Optimal path and gait generations simultaneously of a six legged robot using a GA-fuzzy approach. Robotics and Autonomous Systems, 41: 1-20.